REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	8.Feb.99	DISSERTATION

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
PSYCHOPHYSIOLOGICAL MEASURES FOR HUMAN ATTENTION LAPSES DURING SIMULATED AIRCRAFT OPERATIONS	
6. AUTHOR(S)	
MAJ CALLAN DANIEL J	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
PENNSLYVANIA STATE UNIVERSITY	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
THE DEPARTMENT OF THE AIR FORCE AFIT/CIA, BLDG 125 2950 P STREET WPAFB OH 45433	FY99-76

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Unlimited distribution In Accordance With AFI 35-205/AFIT Sup 1	

13. ABSTRACT *(Maximum 200 words)*

19990219 132

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

The Pennsylvania State University

The Graduate School

College of Engineering

PSYCHOPHYSIOLOGICAL MEASURES FOR HUMAN ATTENTION LAPSES DURING SIMULATED AIRCRAFT OPERATIONS

A Thesis in

Industrial Engineering

by

Daniel J. Callan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 1998

We approve the thesis of Daniel J. Callan

Date of Signature

Joseph H. Goldberg
Associate Professor of Industrial Engineering
Thesis Advisor
Chair of the Committee

Andris Freivalds
Professor of Industrial Engineering

M. Jeya Chandra
Professor of Industrial Engineering

William J. Ray
Professor of Psychology

James R. Comstock, Jr.
NASA Research Scientist
Special Signatory

A. Ravindran
Professor of Industrial Engineering
Head of Department of Industrial and Manufacturing Engineering

ABSTRACT

This study produced a range of aviation performance to which psychophysiological measures were correlated predicting performance decrements due to task overload and vigilance decrement. A high fidelity simulation of an instrument flight pattern produced multiple workload levels resulting in various levels of performance. Psychophysiological parameters including eye movements, EEG, and peripheral temperature were measured. Workload was varied and a secondary task was added to create realistic operational performance levels. Four groups of four subjects provided 64 data segments each during two, 2 hour simulation periods. Eight subjects were instrument rated and eight unrated. Eight subjects had commercial flight experience and eight had no commercial flight experience. Operationally relevant performance levels were based upon Air Traffic Control (ATC) and safety standards. Subjects' performance error was dangerous for 18 of 1024 segments and exceeded ATC standards on additional 193 segments. The Long Fixation parameter was sensitive enough to predict 83% of segments exceeding ATC performance error standards with a 15% false alarm rate.

Factors of workload, attentiveness, and cognitive processing capability affect performance; different psychophysiological parameters are needed to completely describe performance. Level of arousal reflected the "level of attention" for perception, processing, and response execution. The two best arousal parameters, Peripheral Temperature Change and Pupil Diameter Change, were the best performance predictors, these parameters reflected performance decrements related to workload and other stressors. Performance decrements associated with nominal or low workloads were not detected. Saccade Time, Dual Fixation Gate, and seven other parameters related to task

type showed great promise in providing real time feedback on workload levels and the type of task on which operators are engaged.

Elements of cognitive performance were described by the Long Fixation and Short Fixation parameters. A high frequency of Long Fixations was indicative of problem solving activity. A high frequency of Short Fixations was indicative of efficient processing. However, the efficiency was not related to only to workload since subjects used large numbers of short fixations when monitoring the simulation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Advanced Civil Transport Simulator (ACTS): a simulator designed with a two place (pilot/copilot) flight deck with a forward looking out the window graphical interface provided by a silicon graphics Onyx. Flight deck accommodations are similar to those found on a MD-11 or B-777 aircraft.

Airborne Caution and Warning System (ACAWS): the center display panel providing cautions, warnings, and checklists.

Aircraft System Display: on board displays with engine information on the top half and optional system information on the bottom.

Air Traffic Control (ATC): The agency responsible for aircraft separation and sequencing.

Alpha components: EEG rhythms of 9 - 13 Hz, present when an individual is in an alert, relaxed state.

Beta components: EEG rhythms of 14 - 30 Hz present when individuals are excited or highly aroused.

Controlled Flight Into the Terrain (CFIT): an aircraft accident in which a controllable aircraft was flown into the terrain.

Control instruments: altitude indicator and engine power displays providing measurement of aircraft control parameters.

Crew response and evaluation window (CREW): the computer, software and information feeds which computed the Index of Engagement.

Delta rhythm: slow EEG rhythms of less than 4 Hz which are present only during sleep

High Load: the workload factor level in which subjects were constantly required to manually alter the simulator position or velocity.

Horizontal Situation Indicator (HSI): a flight deck instrument display providing a two dimensional overhead view of aircraft position

Index of Engagement: the ratio of the strength of beta brain waves to alpha and theta brain waves.

Instrument Landing System (ILS): a landing approach in which the pilot is provided with glidepath and runway alignment information. The pilot maneuvers the aircraft to

maintain the system indicators on the desired position while landing.

Instrument Meteorological Conditions (IMC): the conditions in which adverse lighting, visible moisture, smoke, dust or a combination thereof, obscure the aircraft/ simulator flight path.

Monitor: the workload level factor in which the subjects monitored the autopilot or co-pilot in control of the simulator.

Mission-Oriented-Terminal-Area-Simulation Facility (MOTAS): the facility at Langley NASA which provided the hardware for monitoring, communication and direction of the simulation from an ATC standpoint.

National Transportation Safety Board (NTSB): the federal agency responsible for air traffic safety and crash investigation.

Nominal: the workload factor level in which subjects manually controlled the simulator in a normal aviation task. Constant corrections were not required.

Performance instruments: the airspeed indicator, altimeter, turn and bank indicator, etc. providing aircraft performance measures.

Precision Approach from Radar (PAR): a landing approach in which a controller from the ground, uses a radar to determine the position of the aircraft relative to the runway and directs the pilot to the desired glidepath and runway alignment. The controller talks the pilot to the ground with a series of turns and descent rates.

Primary Task: the aviator was expected to maintain altitude, speed and course for the simulator flight during this study.

Pupil Diameter: in this study, it was measured by the number of pixels on an horizontal Line across the subject's pupil at the oculometer camera interface.

Reaction Time: the time necessary to respond to normal working tasks (not emergency situations). Reaction time was measured from the initial perception of stimuli to the response.

Saccade: voluntary or reflexive eye movements which are short in duration (20 - 100 msec) and have a relatively high peak velocity between 20 - 600° per second.

Secondary Task: the aviators were given arithmetic calculations to perform during the simulator as a planned distraction.

Simulation Session One (SS-1): The morning simulation session which consisted of 19 segments composed of seven tasks.

Simulation Session Two (SS-2): the afternoon simulation session which consisted of 20 segments with seven tasks.

Task overload: the task load corresponding to the performance decrement on the left side of the Yerkes and Dodson (1908) Stress vs. Performance Curve, Fig. 2.2, p. 10.

Task underload: the task load corresponding to the performance decrement on the right side of the Yerkes and Dodson (1908) Stress vs. Performance Curve, Fig. 2.2, p.10.

Theta components: EEG rhythms waves between 4 - 8 Hz, indicating a state of drowsiness.

Vertical Velocity Indicator (VVI): displays the aircraft's vertical velocity and vertical velocity rate of change information to the pilot.

Workload: the total effort required to correct and maintain the simulation on the ATC directed course, altitude, and airspeed while simultaneously performing any other required tasks.

ACKNOWLEDGEMENTS

Praise the Lord for the gifts he has provided that I could dedicate myself to this project in service to my country and fellow man. The greatest gifts provided were my family and the community that have surrounded and supported me. Without their (less than gentle) prodding this would not have been completed. Especially noteworthy members of that community include my wife, Patti, my under standing children, Bernard, Bridget, Anna, Mark, and Kate, and my advisor Dr. Goldberg.

The support of management and help of technicians at NASA Langley made the effort possible. The patience and advice of my fellow scientists and operators at Penn State and in the US Air Force and NASA made this effort useful to the world.

CHAPTER 1.

INTRODUCTION

The National Transportation Safety Board (NTSB) reported that 50% of all aviation accidents occur during the 4% of flight time comprised of the approach and landing phases of flight (NTSB, 1996). This group of accidents accounts for half of the fatalities from aviation accidents as well. Forty percent of these accidents involved airworthy aircraft flown into the ground by the aircraft operator.

A history of Controlled Flight Into the Terrain (CFIT) accidents by Weiner (1977) documented numerous aviation accidents of this type. Three of these crashes occurred inside of a two year period beginning June 14, 1973 (Weiner, 1977). Although numerous changes in Air Traffic Control (ATC) procedures occurred as a direct result of those three crashes, problems persist. For example, AA Flight 965 flew into a mountainside near Cali, Columbia on December 20, 1995. These CFIT accidents are a result of pilots' inability to attend to the proper cues in the cockpit. In these accidents, the information necessary to avert the accident was available in the cockpit, but was not used due to distraction or absorption (cognitive tunneling) on the wrong indicator. It is important to note, cockpit recordings indicate the pilots were very aroused in the minutes prior to these accidents. The pilots were busy attending to the wrong tasks.

Statistics are not available to describe episodes of inattentiveness occurring in other phases of flight, such as the cruise phase at high altitude. However, commercial pilots flying on autopilot have long complained about drowsiness, boredom, and a general inability to focus their attention on important tasks prior to descent and landing (Fitts and Jones, 1950). This problem is exacerbated today by sophisticated autopilots which can

perform most piloting tasks independently. Fortunately, accidents at cruise altitude are less likely to occur since there are fewer solid objects to hit at cruise altitude. Also, the automatic flight control systems tend to keep aircraft on the correct airspeed, altitude and course, if the autopilot is properly programmed and engaged.

However, even at cruise altitudes, attention lapses can be deadly. Errors committed at cruise altitude, during a low state of arousal, can start a deadly sequence of events. This scenario occurred in the crash of the Boeing 757 near Cali, Columbia (Simmon, 1997). The captain of AA Flt 965 entered incorrect information into the Boeing 757's flight director system, while at cruise altitude. The captain was not in control of the aircraft at the beginning of descent, nor at the time of the crash. He was in task underload monitoring the copilot. Comments made less than one minute prior to the crash indicate the captain was quite relaxed and not attending to the aircraft position information displayed in front of him. Meanwhile, the copilot was overloaded because the runway on which the aircraft was to land was changed.

The majority of CFIT accidents involve task overload. However, the potential cascade effects from task underload are not well understood. As a result of operational complaints of boredom and drowsiness, numerous simulation studies have been undertaken. Attention decrements have been well documented in simulation environments using realistic displays; Pope and Bogart, (1992); Akerstedt, Torsvall, and Gillberg, (1987); Harris, Glover, and Spady, (1986) and Weiner, (1977), documented instances of task overload. However, psychophysiological manifestations of aviation task underload and overload have not been linked directly to performance .

The goal of this study was to establish an analytical relationship between psychophysiological measures and aviation performance. This goal underlay the objectives presented below.

1.1 Objectives

Objective One: Measure aviation performance as related to normal workload, task overload, and task underload

Objective Two: Determine if the psychophysiological measures of eye movement, peripheral skin temperature (arousal), and Index of Engagement, related to aviation workload, and if so, the relationship of aviation workload to those measures.

Objective Three: Determine if the above psychophysiological parameters are related to aviation performance measures, and if so, identify the measure(s) associating poor performance with hazardous states of attentiveness.

CHAPTER 2.

BACKGROUND

As trained professionals, airline pilots are subjected to a certification process requiring safety awareness. This certification process, coupled with the availability of checklists and simulations for practicing emergency procedures, eliminate many errors. In addition, aviators learn specific scanning techniques to provide optimal information processing. If the vehicles these professionals operate are mechanically sound, and if the operators have been certified to possess safe operating skills, what causes accidents?

2.1. Failure to Perceive

First, the operator may fail to perceive the necessary information despite the availability of the information and training designed to prevent such lapses. In a two pilot aircraft, the person at the controls has the sole responsibility of flying. The crewmember at the flight controls is not permitted to do anything but fly or monitor the autopilot as it controls the aircraft. The other crewmember takes on the responsibilities of navigation and communication with Air Traffic Control (ATC).

The flight computer used by the captain of AA Flt 965 to Cali, Columbia displayed information indicating the aircraft had passed a reporting point. However, the captain neither noticed the information nor cross-checked the aircraft position by other available means. ATC had requested a position report when the aircraft was overhead the reporting point but no report was made. The captain set the reporting point, which was behind them, into the computer as a "Fly To" point. Dutifully, the autopilot obeyed and began to turn the aircraft around toward the point they had previously passed. Meanwhile

the copilot reviewed procedures for approaching a new runway. Neither aircrew member had focused his attention on displays for which he was clearly responsible. The captain did not perceive the aircraft position and the copilot did not perceive the aircraft turning.

2.2. Failure to Respond

If aviators do perceive the information necessary to prevent an accident, another type of error may occur. Operators may perceive the information necessary to prevent the accident, but choose not to respond to the information. The copilot flying the aircraft on AA Flt 965 knew the minimum safe altitude for terrain clearance. He was acutely aware of his aircraft's altitude, since he was attempting a rapid descent after accepting a change in the runway to which he was flying an approach. However, the copilot never connected these two pieces of information as an aviator normally would. He was too busy to respond to the information he had perceived.

In accepting a change in the approach he was to fly, the copilot accepted a heavy workload. He focused on learning the new approach. The copilot allowed the autopilot (incorrectly programmed by the captain) to maintain aircraft position and dropped his level of attentiveness to that information. The copilot did not completely ignore the position information, as evidenced by his later confusion about the aircraft's position and heading. However, the copilot did not focus on the information enough to consider the option of climbing, when he discovered his disorientation. His focus was on losing altitude rapidly enough to complete the new approach he was attempting (NTSB Report , 1996).

On the other hand, the captain was so complacent that he suggested they, "Press on," when the copilot expressed confusion about the aircraft's heading and position. Despite having flown in this area previously, he did not maintain an accurate mental model of aircraft position relative to the points he was entering in the flight director. One could argue that what he did was procedurally correct, although his actions made no sense given the aircraft's geographic position. The perceived workload in the cockpit was extremely unbalanced with the copilot unable to assimilate necessary information because he felt overloaded (Siminov et al, 1977) , and the captain failing to perceive his position because he felt underloaded (Comstock, 1987). The workload stress, not the actual workload, plays a large role in the ability to react to that workload (Yerkes and Dodson, 1908; Easterbrook, 1959; Wickens, 1992).

2.3. Failure to Act Appropriately

Finally, after aviators perceive a problem and consider their situation they must take action. Aviators are taught to:

1. Maintain aircraft control,
2. Analyze the situation, and
3. Take appropriate action.

Failures can occur in analyzing the situation, selecting the appropriate action, or executing the appropriate action. For example, on AA Flt 965: (1) The copilot analyzed the situation and determined there was a problem due to disorientation. (2) His response should have been to climb to an altitude higher than surrounding terrain by initiating an emergency climb with maximum power, while minimizing aerodynamic drag. The

aircrew selected maximum power just prior to impact, and never brought in the speed brake.

2.4. Human Information Processing Model

The model (Figure 2.1) of Human Information Processing (HIP ; Wickens, 1992, p. 17) implies that the limiting factors in a stimulus-response closed loop are attention resources (Posner, 1980) and the number of channels available. This model can account for the failures previously described.

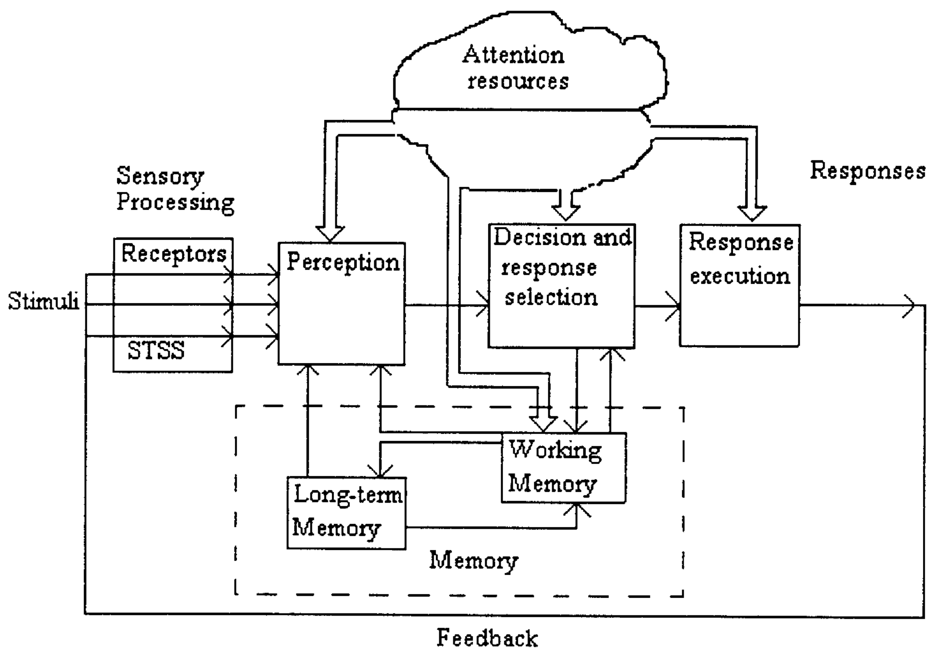

Figure 2.1. Human Information Processing Model.

2.4.1. Channel Capacity as a Limiting Factor

It is possible that channel capacity acts as a limiting factor in aviation tasks, especially in the case of novices. However, numerous studies (Wickens, Bellenkes, and Kramer, 1995; Mourant and Rockwell, 1972; Kopp and Liebig, 1990) have demonstrated novices can perform the basic instrument cross-check and control tasks. In the worst case, one could assume novices perform an instrument cross-check in a serial manner using a single visual perceptual channel. Since the task could be performed by subjects in all studies above, the single channel limitation does not preclude use of the HIP Model. Thus, the HIP Model may be used to analyze cognitive issues related to all levels of aviation expertise.

2.4.2. Attention Resources as a Limiting Factor

Studies of skilled activity have demonstrated attention resources are a limiting factor in performance of the skilled activity (Broadbent, 1977; Posner, 1980). Several definitions are necessary to understand how attention plays this limiting role (Wickens, 1992, 74-77).

Focused Attention - The ability to perceive, process, and respond to a desired stimulus. Ability to focus attention is reduced by perceptual competition, and display clutter, but it may be enhanced by display redundancy.

Divided Attention - The ability to simultaneously perceive, process, and/or respond to more than one stimulus. For example, an aviator may be responding to a previously perceived need to turn an aircraft while simultaneously perceiving an error in airspeed.

Selective Attention - Used in situations in which a person is required to sample multiple sources of information periodically. The sample may be perceived using either focused or divided attention. The sampling strategy is considered optimal when the scan pattern's expected value is optimized, or expected cost is minimized.

2.5. The Attention Spotlight

Attention has been compared to a spotlight (Wachtel, 1967). Availability of attention resources affects the HIP model at both extremes. Excessive availability of attention allows perception of extraneous information (distractions; Posner, 1980). In this case, the information spotlight is too broad to focus on the desired stimulus alone. A reduction in attention resources narrows the spotlight. Spotlight size is optimal when wide enough to allow perception and processing of necessary information, but narrow enough to prevent collection of extraneous data. The spotlight is too narrow if it does not include all necessary information.

Although the concept of an attention spotlight was unknown at the turn of the century, factors affecting the spotlight were first explained around this time by Yerkes and Dodson (1908). Yerkes and Dodson linked arousal to the level of performance in rats. This link of performance with arousal was subsequently explained in terms of attention resources in humans (Easterbrook, 1959).

Initially, the attention spotlight is large in a relaxed, but conscious person. As the human is aroused or stressed, the attention spotlight begins to narrow. The narrowing of the spotlight helps focus attention on the required task. Eventually, an optimum level of stress produces peak performance for the given task (Easterbrook,1959).

The optimum level of stress varies with each task and the person's level of competence. Increasing the stress beyond the optimal point narrows the focus of attention to a point where important information is excluded and performance degrades (Figure 2.2.).

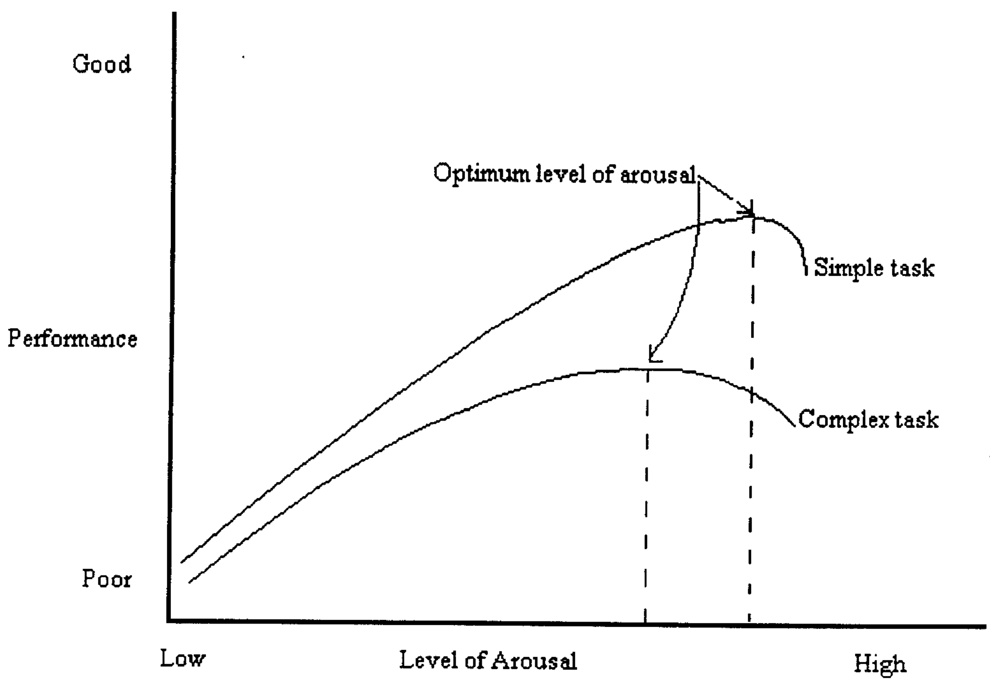

Figure 2.2. The Yerkes-Dodson Law

2.6. Cause of Hazardous States of Attention

Hazardous states of attention occur whenever the demands of a task exceed the attention resources available. Tasks can be safely accomplished when attention resources are above some minimum level and they are best accomplished when attention resources are at an optimum level (Figure 2.2.). The level of arousal or relaxation determines the attention resources available. Level of arousal is likewise related to the amount of

cognitive activity (Pope and Bogart, 1992; Prinzel, Hitt, Scerbo, and Freeman, 1995). The nature of this relationship will be discussed below.

Since level of arousal affects attention resources, it also affects the probability of encountering various hazardous states of awareness. Absorption occurs at high levels of arousal when the attention spotlight shrinks too much (Easterbrook, 1959). High levels of arousal would correspond to the downward slope on the right of the arousal-performance curve (Figure 2.2.). An individual who is bored will seek sensory stimulation and consequently create a broad attention spotlight. Low levels of arousal would correspond to the upward slope on the left of the arousal-performance curve. Two other hazardous states of awareness, sleep onset (Nicholson et al, 1989; Makeig and Inlow, 1993) and distraction (Wickens, 1992, p. 74), occur at any level of arousal, but they occur more frequently at low levels of arousal. Distraction is most prevalent when there has been a broadening of the spotlight.

2.7 The Instrument Cross-check Task

How do hazardous states of attentiveness affect performance of aviation tasks? The primary task of the pilot controlling an aircraft is cross-checking of cockpit instruments to determine aircraft position, velocity, and acceleration.

Aviation instruments are divided into two broad subcategories, control instruments and performance instruments. Control instruments provide feedback on control inputs. In commercial airliners control inputs are made via engine throttle(s) and control yoke/control stick. The control instruments providing feedback for these inputs are the Engine Pressure Ratio (EPR) and attitude indicator respectively.

When flying under Instrument Meteorological Condition (IMC) flight rules, pilots must file a flight plan specifying their intended route of flight, airspeed(s) and altitude(s). Unless otherwise cleared by Air Traffic Control (ATC), the pilot is expected to adhere to the flight plan. ATC instructions supersede filed flight plans. Performance instruments provide feedback on aircraft position, velocity, and acceleration. ATC is not concerned with control instruments, but if performance instruments should deviate from the norm, ATC should intervene rapidly.

To maintain the flight parameters specified by either flight plan or ATC a pilot must constantly check instruments to ensure aircraft parameters comply with requirements. If the aircraft is off conditions the pilot must obtain several pieces of information to determine the action to be taken. For example, the pilot may determine the aircraft is below its assigned altitude by looking at the altimeter. The next piece of information desired is the aircraft's vertical velocity and the vertical velocity's rate of change. These pieces of information are obtained visually by fixating the Vertical Velocity Indicator (VVI) to determine a current value, comparing the value to a value on a legend, and then dwelling on the indicator to identify a trend. Sometimes it is also useful to dwell even longer on the VVI indicator to determine the rate of acceleration.

If the aircraft was low and the vertical velocity was upward at a stable rate, the pilot may choose to take no action. However, if the vertical velocity was downward and accelerating, the pilot would probably choose to make a control input to reverse the trends. A control input requires the following steps summarized from AFM 11-217 (1996):

1. Determine current pitch attitude (reference point).
2. Pull back on the control yoke/control stick to initiate pitch attitude change.
3. Track the pitch attitude change from reference point until reaching lead point to reverse back pressure.
4. Push forward on control yoke/control stick to establish new pitch attitude reference point.
5. Determine current pitch attitude as compared to desired target attitude.
6. Cross reference VVI to determine if the new pitch attitude has achieved the desired effect on vertical velocity (if not, return to step 1).
7. Cross reference altimeter to estimate the required duration of the correction.
8. Track altitude change until reaching lead point to initiate control input to take out correction.

If each step is performed correctly, a single altitude correction requires eight steps and use of three instruments, the attitude indicator, the altimeter, and the VVI. Failure to perform any step of the process correctly, requires insertion of corrective steps.

The narrative above describes steps necessary to correct one parameter individually. Normally, a pilot is monitoring conditions of airspeed, altitude, and course simultaneously. Thus, the term cross-check applies both to the perceptual pattern among the parameters to be monitored and within the process of affecting one parameter.

2.8. Hazardous States of Attention

Both Single-Resource Theory (Kahneman, 1973) and Multiple-Resource Theory (Navon and Gopher, 1979) acknowledge the role of arousal in supporting the pool(s) of attention required to perceive, process, and respond to stimuli (Wickens, 1992). Hazardous states of attention occur when the attention searchlight is narrowed by either too much or too little arousal (Yerkes and Dodson, 1908).

2.8.1. Absorption

Absorption occurs when the attention searchlight is narrowed by arousal. The mind's attention resources are completely occupied in a non-optimal manner. Actual workload may not be excessive, but perceived underload or overload causes misapplication of attention. Absorption is characterized by too much attention to one area of interest, preventing optimal allocation of attention resources to other areas when divided attention is necessary. It is also characterized by high levels of cognitive activity (Prinzel et al, 1995). Although attention is devoted to an important area of interest, dwell time on the particular area of interest is too long, or dwells occur too often to allow for optimal scanning strategy incorporating other areas.

2.8.2. Distraction

A distraction is something that diverts attention away from the desired focus of attention. Unlike absorption, when the operator is looking at an appropriate location too long, distraction demands attention be devoted to areas outside the primary occupational focus. It is a failure of focused attention. The power of the distracter determines its effectiveness in grabbing the attention spotlight. The more focused the spotlight, the more difficult it is for a distracter to grab attention.

2.8.3. Vigilance Decrement

Vigilance decrement is associated with an increase in the response criterion. The criterion shift results from a decrease in the frequency of target events in repetitious, or 0tedious work (Wickens, 1992). It occurs at a low state of arousal and is sometimes a

precursor to sleep (Ogilvie et al, 1988; Hori, 1982). Therefore, within the context of an aviator's cross-check, vigilance decrement would occur when the cross-check task offers no new, or unique information. For example, when systems are operating normally for extended periods of time the cross-check becomes repetitious and tedious; boredom commonly ensues in these situations (Makeig and Inlow, 1992; Mackworth, 1948). Vigilance decrement manifests itself as a criterion shift which may be reflected in psychophysiological measures to be discussed later.

The progressive changes from wakefullness to sleep are best characterized by the convergence of a number of indices (Ogilvie et al, 1988). These indices include brain wave activity, heart rate, and breathing pattern. However, if the intent is to prevent an operator from entering a near sleep state convergence of these indices would be too late since some indices do not change until after normal eye movement has ceased. This study will not include the topic of sleep onset, only vigilance decrement.

2.9. Pyschophysiological Measurement

Numerous factors in choosing psychophysiological measurements have been employed to evaluate the mental states of vehicle operators. Candidates for measurement include cardiac activity, peripheral vascular activity, skin conductance, electroencephalography (EEG), pupillography, oculomotor activity, body movements, and others (Stern, 1987). Before selecting a method of evaluation several factors must be considered. What is the target group? What are the mental states the method is attempting to describe, and in what working environment? Is it reasonable to employ the chosen measurement technique in the given operational environment? Will the

technique, alone, provide enough descriptive information to quantitatively describe the operator's ability to perform in the operating environment?

2.9.1. Target Group

This study assumes aviators will be required to continuously operate vehicles for periods of time exceeding one hour. The motivation for this study is concern that people operating vehicles in these environments, may endanger themselves and others through a failure to attend to information necessary to operate their vehicles in a safe manner. The design of experiment acknowledges aviators have different levels of proficiency.

2.9.2. Working Environment

Aviators are required to perform multiple tasks in a dynamic environment. A pilot must simultaneously monitor airspeed, altitude, heading, vertical velocity, etc.. Unlike situations in which a laborer must devote full attention to a single task, vehicle operators must attend to multiple tasks, dividing attention among the tasks. Many distractions are built into the aviation environment. For this reason, two pilots often share the crew duties explained earlier.

In aviation the rule is aviate, navigate, and communicate. The primary task is to maintain aircraft control, the secondary task is to navigate, and the tertiary task is to communicate with ATC. All three are required to safely complete the overall task.

At a lower level, the cross-check used to accomplish the overall objective is prioritized first to maintain control. Since power can be set to a required level and left static, the attitude indicator becomes the center of the pilot's cross-check. Next,

performance instruments such as the airspeed indicator, altimeter, vertical velocity indicator, and course indicator are required to navigate through three dimensional space. Finally, radios and navigational aids provide instructions for navigation and collision avoidance.

2.9.3. States of Attentiveness

At one end of the attentiveness spectrum, vigilance decrement describes a mental state in which the operator looses the ability to perform the perceptual tasks necessary for safe operation of a vehicle. This state, sometimes described as "Drowsiness," is the irresistible urge to close ones eyes while attempting to perform a task requiring continuous visual input (Nicholson et al., 1989). On the other end of the spectrum, "alertness" is often used to describe the quality of an aroused, attentive mental state.

2.9.4. Physiological Measurements in the Aviation Environment

EEG is a difficult tool to use for identification of mental state in an operational environment. The preparation and processing required to make EEG useful, preclude its use on a day to day basis in an operational environment, although it has been useful on a limited basis (Akerstedt, Torsval, and Gillberg, 1987). A significant body of literature exists relating electroencephalogram (EEG) measurements to some mental states (sleep and drowsiness), but it has not proven effective in identifying other hazardous and alert mental states.

Another, less intrusive, but potentially useful psychophysiological index is peripheral temperature. Peripheral temperature is a useful index, in a comfortable

environment with stable air temperature and humidity. This may not a practical requirement in many working environments, but stable environmental control in aircraft is possible. Data is available relating peripheral temperature to stress (Grimsley, D. L., 1994; van Quekelberghe, R., 1995) and relating stress to arousal (Yerkes and Dodson, 1908; Easterbrook, 1959; Siminov, et al., 1977). However, no database exists relating peripheral temperature to arousal.

Pupillography and measures of oculomotor activity provide a non-intrusive means of gathering physiological measurements in an operational environment. Remotely mounted eye trackers can measure information about subjects' saccades, fixations, blinks, slow eye movements, and pupil size. However, eye tracking has not been demonstrated to provide sufficient resolution to differentiate among various states of productive cognitive activity, and various hazardous states of awareness. If analysis of eye movement does provide sufficient resolution to link it to performance, it would provide an unintrusive tool to monitor aviators.

Cardiac activity, skin conductance, and body movements were considered for use in this study, but were not selected due to the variability between and within subjects.

2.10. Eye Movement Basics

Several methods of recording eye movements exist. Low resolution of eye movement is available from videotape of a subject's face and eyes (Mourant and Rockwell, 1972; Cole and Hughes, 1988). Data reported using these techniques often do not include an estimate of resolution, but instead reports separation of objects viewed by the subjects. Slightly better resolution results from placing electrodes beside the eyes to

measure the changes in electrical potential generated as the poles of the eyeball move. This method, electrooculography (EOG), provides an excellent means of identifying saccadic eye movement, but fixations must be inferred from summation of saccadic movement. A third method, infrared oculometry, involves projecting an infrared light into the eye, and recording the relative positions of the first reflection (off of the cornea), and the first purkingie image while the subject views a calibrated area. Accuracy using this method is normally 1° of visual angle when recorded at 60 Hz., and resolution is 0.5 min arc/sec (Saito, S., 1992).

2.10.1. Saccades

The most prevalent and important type of eye movement is the saccade. Saccades are short in duration (20 - 100 msec), and have a relatively high peak velocity between 20 - 600° per second (Hallett, P., 1986). Saccades may be either voluntary, or reflexive in nature. For example, a saccade to a specific view point may be made in response to instructions, but since saccades often miss the target (Bahill and Stark, 1975), a second, short, reflexive saccade may be necessary to complete the eye movement to the desired position.

Variation in saccadic velocity results largely from the different velocities associated with small and large saccadic movements. Relatively small movements between targets close together result in low saccadic velocities, whereas long movements incorporate higher velocities.

2.10.2. Fixations

A fixation can last as little as 70 msec, as long as 400 msec, or longer. These extremes are low probability situations. On the average, a highly motivated scan pattern will possess three fixations per second (Boff and Lincoln, 1988). Saccadic movement accompanying fixations will require approximately 33 msec each, meaning each fixation averages about 300 msec. However, this rapid scanning applies to a free scanning paradigm, and assumes little or no processing time is required for the scanned scene.

Fixations requiring some amount of cognitive processing slow down the perception process (Saito, 1992). For example, studies were undertaken to quantify the changes occurring in an instrument cross-check when a new digital radar altimeter replaced an old analog altimeter (Harris and Glover, 1985). Results showed the experienced pilots' cross-checks were slowed by the new digital altimeter. Interestingly, the increase in time did not occur on the new (more difficult) instrument, but on the subsequent fixation where the subjects were apparently processing the new data format. Other aviation studies (e.g., Wickens, 1994) have shown similar results with degree of difficulty, and fixation times. However, less experienced pilots tend to increase fixation time on the instrument providing the difficulty. Fixation time increases with cognitive workload, but workload is not the only factor increasing fixation time.

John Stern (1987) suggested that oculomotor activity may be a useful indicator of fatigue, or alertness since saccadic velocity and frequency were lower after a number of hours on task. Although this reference to fatigue was somewhat ambiguous, the issue was clarified by later research. Oculomotor muscles do not fatigue; saccades do not slow due to fatigue to the optical musculature (Saito, 1992). Average saccadic velocity is

reduced by execution of misplanned saccades called glissades (Bahill and Stark, 1975). This misplanning indicates the saccadic planning capacity, not the muscles suffer from a form of mental fatigue, or attention deficit. For the same task, fixation duration increased over the course of the five hour study, indicating cognitive processing was slowing. Thus, eye movements from the same instrument cross-check performed in an alert state of attentiveness versus a hazardous state of attentiveness, may result in different fixation durations and saccadic profiles.

2.10.3. Pupil Diameter

Pupil diameter generally increases as alertness increases, and decreases as alertness decreases. However, pupil function is a very complex mixture of voluntary, and reflex functions (Stern, 1987; Gray, 1977). Vergence, focus, and light reflexes all affect pupillary function. Lighting, distance to the display, and display resolution must be held constant to make pupil diameter a useful measure.

2.10.4. Blink Frequency and Duration

The eye blink is a very opportunistic mechanism required to cleanse and moisten the cornea's surface (Gray, 1977; Skelly, 1993). When visual perception is not required or higher order control mechanisms perceive that active visual scanning is not required, the blink rate and duration increase. In vigilance tasks, there is an increase in the frequency and duration of blinks with time on task (Stern, 1987). Blinks often occur in conjunction with saccadic movement, since perception during saccadic movement is

either extremely limited or nonexistent. A breakdown of this saccadic/blink coordination seems to occur with loss of attention (Skelly, 1993)

2.11. Electroencephalogram (EEG) Measures as related to Attention

The body of literature relating EEG to attentiveness leads to two conclusions. First, EEG has been used extensively to define sleep, but little work has been done to characterize states of attentiveness. Second, any attempt to characterize attentiveness with EEG requires use of several components of the EEG frequency spectrum (Okogbaa, 1994; Makeig and Inlow, 1992; Stern, 1987; Akerstedt, Torsvall and Gillberg, 1987). The EEG frequency spectrum is characterized in Table 2.1.

Table 2.1 EEG Frequency Breakdown

Frequency Band Designation	Cognitive State	Frequency Range (Hz)
Delta	slow waves, sleep only	$f_0 < 4$
Theta	indicate drowsiness	$4 \leq f_0 \leq 8$
Alpha	relaxed alert	$8 \leq f_0 \leq 13$
Beta	high frequency, cognitive	$13 \leq f_0$

Grandjean (1981) described the above EEG frequency components as follows.

Delta rhythm. Delta (less than 4 Hz) components, like the theta components, are slow waves and are present only during sleep.

Theta components. Theta (4 - 8 Hz) rhythms indicate a state of drowsiness. They replace the alpha components at the onset of sleep.

Alpha components. The alpha rhythms include an electrical activity with frequencies of 9 - 13 Hz. Alpha rhythms are present when an individual is in an alert relaxed state.

Beta components. Beta components (14 - 30 Hz) are associated with states of excitement or arousal. The presence of high components of beta rhythms is manifested in the form of increased alertness.

The above description provides a concise and comprehensive understanding of EEG components. It appears an aviator's state of alertness could be completely described using Grandjean's definitions. However, some practical difficulties exist in accepting the above definitions for the vehicle operator. First, the clinical definition of sleep used above indicates the many drivers who "fell asleep at the wheel" of their vehicle actually did not fall asleep. These drivers were only "drowsy/relaxed at the wheel."

Second, if theta components replace alpha rhythms at sleep onset, and alpha rhythms indicate an alert, relaxed state, then there is no room for drowsiness and the process of falling asleep. Finally, what are the differences among relaxed alert (alpha), alert (beta), and increased alertness (high beta), with respect to actual states of attentiveness, and what characterizes the transitions among these states?

Fortunately, Grandjean (1981) also caveats his statements as general, and lacking in explanation of transitions. In fact, Grandjean and many others following him (Okogbaa, 1994; Makeig and Inlow, 1992; Stern, 1987; Akerstedt, Torsvil and Gillberg, 1987) agree that some method of integrating the different EEG components may be best to judge alertness and the transitions among various states of attentiveness.

2.11.1. Nominal EEG rhythms and transitions

EEG rhythms, like most of nature, follow multiple embedded cycles. In the human, the circadian rhythm describes one of the long period cycles that affect the human. This rhythm includes the daily sleep/wake cycle as well as natural declines and rises in human performance through the day (Astrand and Rodahl, 1986). Embedded within the long circadian rhythm are shorter rhythms, only two to three minutes long (Simonov, 1987).

Other EEG based studies (Makeig and Inlow, 1993; Akerstedt, Torsvall, Gillberg, 1987) report subjects drifting in and out of specific mental states. However, some studies note that mental state transitions were rapid and irreversible (Okogbaa, Shell and Filipusic, 1994; Nicholson et al, 1989), more like an exponential function rather than a sine wave. However, the type of transition (sine wave vs exponential) does not affect the performance characteristics of the transition states. Daytime sleep latencies vary considerably, but have no bearing on the relationship between EEG and performance (Nicholson et al, 1989). Thus, transitions among EEG rhythms and among related mental (performance) states are independent of time in the previous state. Furthermore, Nicholson's study demonstrates transitions from multiple daytime mental states into a drowsy state.

2.11.2. Index of Engagement (IE)

Experts have noted the need to consider the different types of EEG activity, and transitions among states, to accurately appraise mental state. Index of Engagement (IE) (Prinzel III et al, 1995b; Prinzell III et al, 1995a; Pope, Comstock, Bartolome, Bogart,

and Burdette, In Press) is a ratio of the strength of high frequency brain waves (Beta), over the strength of low frequency brain waves (alpha and theta). Strength is determined from power spectrum density of the brain waves from the central parietal (Pz) site. Beta rhythms are strongest when there is intense cognitive activity. Alpha components grow in strength as a subject enters a "relaxed" mental state, and give way to theta components at sleep onset (Okogbaa et al, 1994). Thus, the IE provides an analysis tool which combines three primary components of the EEG.

Index of Engagement can be calculated from any snapshot of EEG, and is easily incorporated into feedback loops to modify the degree of difficulty of a given task (Prinzel III et al, 1995a; Prinzell III et al, 1995b; Pope et al, In Press). Both positive feedback (Prinzell III, et al, 1995a), and negative feedback(Pope, et al, In Press; Prinzell III, et al, 1995a) loops have been implemented using IE feedback. The task used in these studies was the Multi Attribute Task (MAT) which provides a task having attributes similar to those of the aviation cross-check, however the task can be performed on an 286, or later personal computer. Positive feedback of the Index of Engagement drove the system unstable. On the other hand, negative feedback demonstrated the ability to stabilize the level of engagement desired, if the gain was high enough to require some constant level of attention by the subject (Prinzell III, L. J., Scerbo, M. W., et al, 1995). The real time feedback provided by IE provides a straightforward use of EEG to incorporate EEG into design of experiment.

2.12. Peripheral Body Temperature as Related to Mental State

Skin temperature varies as a function of stress and exertion (Takenaka and Zaichkowsky, 1990; Astrand and Rodahl, 1986; Pergola et al., 1994). The amount of blood flow to the peripheral capillaries affects these changes. Vaso-motor responses, which control blood flow to the skin, can be divided into two responses, vaso-constriction and vaso-dilation.

As stress increases, cardiovascular output remains constant, but blood flow is redistributed. Stress causes peripheral skin capillaries to vaso-constrict, while at the same time, skeletal muscle capillaries vaso-dilate. These vaso-motor responses provide a rich supply of blood to the muscles in preparation for muscular exertion in response to perceived stress (fight or flight response). Normally, physical exertion follows this initial response, and brings a vaso-dilation response to the peripheral skin capillaries to cool the body. When stress occurs in a white collar work environment, physical exertion does not follow and peripheral temperature affected by vaso-constriction, or lack thereof, can act as an index to perceived work stress.

Peripheral blood flow, and thus peripheral skin temperature, can increase through two different mechanisms. First, vasodilatation occurs in conjunction with physical exertion in order to increase the blood supply to the skin surface and allow for loss of body heat generated in conjunction with physical exertion. Second, if stress is abated, the blood supply to the skin also increases, but this is due to a lack of vaso-constriction, not due to vaso-dilation (Pergola et al., 1994). Thus, when the human body is in a state of low physical exertion, vaso-dilation is not a factor, and vaso-constriction not only controls blood flow to the skin but peripheral skin temperature also.

The predictable response of skin temperature to stress, whether physical or mental, makes it a good candidate to measure stress and arousal. Peripheral skin temperature can provide an index to perceived workload since this is a stress to which vaso-constriction responds. If a person is stressed and no physical exertion takes place, peripheral skin temperature decreases. If a person relaxes, there is an accompanying increase in skin temperature. Skin temperature should remain a constant index of stress level during periods of low physical exertion.

There are several drawbacks to the use of peripheral temperature. First, if physical exertion (i.e. rapid hand movements) or environmental heating causes build up of body heat, vaso-dilation will overcome a vaso-constrictive response to allow cooling of the body. Second, peripheral skin temperature is extremely responsive to biofeedback (Grimsley, 1994; van Quekelberghe, 1995). In fact, very localized changes in skin temperature are possible by thought control (van Quekelberghe, 1995). Third, stress form unknown sources can produce variable results. An experimenter may have difficulty dealing with stress induced by discomfort or illness.

Research environments in which the subjects are not completely engaged by the realism and demands of the task would allow opportunity for biofeedback to affect peripheral temperature. Attempts to use peripheral temperature as a measure of perceived workload should attempt to provide the following conditions:

1) White collar work environment (no extreme physical exertion) ,
2) Stable and comfortable environmental conditions, and
3) Realistic, demanding work environment.

2.13. Cognitive Activity Required in Aviation

Numerous activities ranging from aircraft maintenance to processing of auditory stimuli are required for safe operation of aircraft. However, this study will deal with a small subset of activities accomplished during operation, which require visual input. The types of tasks required for safe airline operations are broken down by type of task.

2.13.1. Reaction Time Task Description

When considering the topics of safe vehicle operation and reaction time, initial focus is usually on time required to physically respond to an emergency situation. This study was concerned with relating task performance and psychophysiological measures to work load. Reaction times considered here refer to the time necessary to respond to normal working tasks (rather than emergency action). Reaction time was measured from stimuli onset to response.

Usually, reaction time is related to two scanning processes. Visual scanning reaction time is a linear function of externally viewed set size terminated upon acquisition of the target (Neisser, 1963), whereas memory scanning reaction time is a linear function of set size of a memorized image based on an exhaustive search. Time for the search depends upon the size of the positive set (Sternberg, 1969; Liu, 1996). The relationship between reaction time and set size is affected by practice (Humphrey, 1994).

Recent developments in learning theory and problem solving have demonstrated the importance of creating strategies (Gopher, 1994), or production procedures (Anderson, 1993) as part of the learning process. The importance of these strategies is

their transference to other similar problems. A novice would not have sufficient time to create these strategies. Thus, the effect of set size would be predictable.

However, it is a well known fact that practice reduces both error and reaction time for many processes. When vehicle operators have developed very efficient cross-checks, they are said to have automaticity. Automaticity is characterized by three major properties: (1) Automatic processing is speeded because it is not limited by available processing resources, (2) Processing is effortless, with respect to cognitive resources, since it requires no cognitive resources, and it suffers no dual task interference for the same reason, and (3) Automatic processing is obligatory (not controlled by resource allocation) and is driven by resource presentation alone (Logan, 1991). A commonly accepted indicator of automaticity is loss of set size (frame size) effect (Healy and Fendrich, 1992).

2.13.2. Memorization Task Description

Operators are required to memorize externally imposed operating criterion such airspeeds. Unlike static mechanical limits placed on aircraft which will result in mechanical damage or failure if exceeded, externally imposed limits are continually changing and do not provide feedback through mechanical failure. For example, as an aircraft transitions from high altitude cruise to landing, airspeed drops gradually from 480 knots to much lower speeds (138 knots in this study). The aviator must memorize, or use a memory aid to provide a reference speed limits to compare to actual airspeed. This task is not typically challenging, requiring one fixation to register the pertinent data (Spady and Harris, 1983).

Memorization is also required for navigation points and route structures. However, the cognitive task is more challenging since the perceived information must be incorporated into a mental model, or referred to multiple times. Building a mental model does not require extensive visual input, but cognitive activity is high (Skelly, 1993; Just and Carpenter, 1976).

2.13.3. Mental Arithmetic Task Description

Typically, aviators calculate fuel and time required from the present position to the destination. These estimates are subsequently added to the present time, or fuel state to create an estimate of arrival conditions. There are a variety of mental arithmetic problems ranging from simple multiplication (2 x 2 = ?) to division/multiplication by fractions. If this is a memorized outcome (2 x 2 = 4) the task would require little cognitive energy, whereas intense problem solving requires greater mental output (*beta* strength; Okogbaa, Shell, and Filipusic, 1994). In aviation, there is typically some mix of these operations occurring while the aviator is sharing attention with other operational requirements. Since this is not part of the primary flying task, it can be considered a planned distraction.

2.13.4. Comparison Task Description

The majority of a vehicle operator's time is spent on comparison tasks. What is the actual speed compared to the desired speed? What is the vehicle position versus the desired route? How does following distance behind the next aircraft compare to a mental

model of what is safe for current environmental conditions? How much time/fuel is required to reach a destination compared to desired time/fuel state?

Comparisons can be made with amazing speed (200 msec/comparison; Just and Carpenter, 1976). Assuming cognitive challenge is proportional to the speed with which a task is accomplished, comparisons would not be cognitively intensive. However, if the comparison involves a mental model, comparison time will increase with the number of anchors required for the model.

2.13.5. Issues with Cognitive Models

Automaticity plays a large role in effective operation of aircraft. Any cognitive model describing attention allocation for vehicle operation must include consideration of automaticity. However, studies have demonstrated that individuals rarely function with complete automaticity. Instead, experienced individuals function at some level close to total automaticity, and less experienced individuals function with lower levels of automaticity (Logan, 1985; Logan, 1991; Healy and Feudrich, 1992).

Two differing theoretical approaches explaining automaticity are Multiple Resource Theory (Wickens, 1992) and Automaticity-as-Memory Theory (Logan, 1991). Both theories fit within a Human Information Processing Model (Resource Theory) but Automaticity-as-Memory Theory affects potential feedback loops and allocation of attention.

Multiple resource theorists argue that practice, which leads to automaticity, simply reduces the level of attention required for a given task. Benefits are derived from reduced requirements for attention resources in the decision and response selection phase, and in

the response execution phase. Excess attentional resources can then be applied to add another response loop within the initial time criteria, or the original loop can run more rapidly.

Memory theorists counter with examples of automaticity which demonstrate tasks requiring no additional resources (Logan, 1985). Furthermore, there is no construct for learning automaticity within the resource model (Logan, 1991). Automaticity-as-memory theories use a power law approach to explain the relationship between practice and automaticity (Newell and Rosenbloom, 1981; Logan, 1991). The form of the equation is:

$$\mathrm{RT} = a + b\mathrm{N}^{-c}$$

where RT is reaction time, a is an irreducible asymptote, and b is the difference between initial and asymptotic performance, N is the amount of practice (expressed as number of trails/sessions, and c is the learning rate (Logan, 1991).

The automaticity-as-memory approach demonstrates a method by which the algorithmic, resource theory, can be circumvented. It does not deny the theoretical process outlined in the original Human Information Processing Model (Figure 2.1), but it would add an additional path through the process like that in Figure 2.3. Furthermore, this theory accounts for the learning of automaticity.

The potential drawback of the automaticity-as-memory theory is its claim to obligatory processing of stimuli. This would insinuate an expert could not fail to respond to any stimuli which are part of automatic processes. However, response failures do occur in hazardous states of attentiveness (Makeig and Inlow, 1993; Nicholson et al, 1989), but it is unclear where the failure takes place. Is the stimulus never perceived, or is it improperly processed? If attention is not required to process obligatory stimuli,

hazardous states of attentiveness should not affect automaticity when the proper stimulus is fixated.

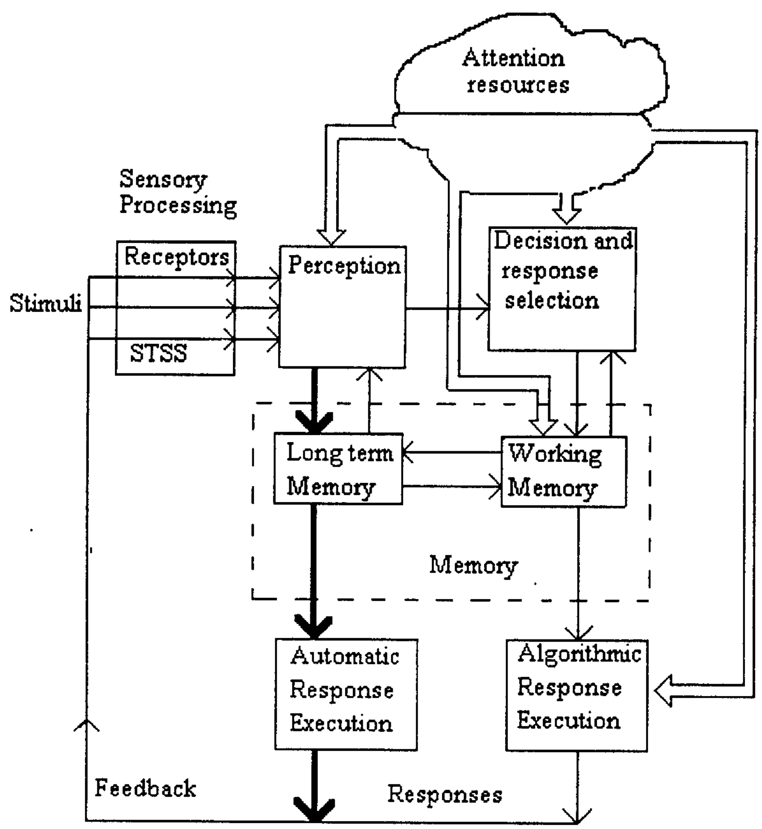

Figure 2.3. Information Processing Model with Automaticity- as-Memory

Finally, there is no accounting for loss of automaticity over time; there is a large range of forgetting rates depending on the task involved (Healy and Feudrich, 1992). Once an aviator learns to operate an aircraft with automaticity, the question remains whether these automatic aircraft cross-check strategies are part of permanent storage (Healy, Fendrich, Crutcher, Wittman, Gesi, Ericsson, and Bourne, 1990). However,

procedural strategies (like the cross-check) are more likely to be permanently stored skills for automaticity (Healy and Fendrich, 1992). These two issues do not contradict automaticity-as-memory theory, but they do question the completeness of the theory, since loss of automaticity is not considered.

If automaticity-as-memory theory fits into the model proposed in Figure 2.3, there are interesting questions about the transitions between the two process loops. Logan (1991) proposes a race model to account for the transition from an "algorithmic" closed loop to an automatic closed loop. This race model presumes the stimulus-response loop is either in an algorithmic state or it is in an automaticity-as-memory state. There is no mix. Given these two distinct states, and assuming the time to complete a stimulus-response loop is appreciably shorter with automaticity, there should be a jump discontinuity in the progression of time to complete a task, as an operator transitions from a practiced, resource limited state, to an automatic state. The same would be true when regressing to an algorithmic state. On the other hand, a resource theory perspective would call for a gradual decrease or increase in response time with practice or regression. The closed loops resulting from automaticity would exhibit shorter reaction times than those without automaticity, in either case.

Distraction, absorption, and vigilance decrement result in increased reaction time. If hazardous states of awareness caused loss of automaticity, this would account for the accompanying increase in reaction time associated with attention deficits. A gradual increase in reaction time would support a multiple resource model in which the attention assets required to accomplish the task gradually decreased to a minimum level where

automatacity occurred. Whereas a jump to and from automatic reaction times would support an automaticity-as-memory theory.

2.14. Eye Movement Relationships to Cognitive Workload and Attention

Several relationships between eye movement and perceived workload, and between eye movement and attention have been developed above. For instance, with a motivated subject, up to three fixations per second may occur in free scanning. However, as cognitive load increases the number of fixations per second decrease. Likewise, cognitive load increases with distraction and absorption, if these situations require cognitive processing for analysis. Other states of attentiveness like vigilance decrement and normal working states generally display peaks and valleys in cognitive workload which correspond to biorhythms.

Any attempt to quantify eye movements related to performance must consider the questions related to the operators' stress level. If relaxed, operators can becomes bored or daydream (internal distraction). These are states related to low stress levels, although not exclusively. Alternatively, high levels of stress due to perceived workload are more likely when operators are engaged in a difficult work environment. In this environment absorption, external distraction, and normal work activity are likely to occur. Thus the first step to quantitatively describe operators' states of attentiveness, is to ascertain their stress level or arousal. Changes in arousal act as an indicator of the states of attentiveness likely to occur (Khaneman, 1973). For example, an operator may stare at a single spot on the control panel for different reasons. If relaxed, the stare could indicate daydreaming, but if stressed, the stare could indicate absorption on an anomalous indicator reading.

The first steps to quantitatively link performance and psychophysiological measures are to consider operator performance and physiological measurements as related to different workload levels. Aircraft operation is a relatively well defined task so it is possible to vary workload level to determine the affect on performance and psychophysiological parameters.

2.14.1. Comparison Task Effects on Psychophysiological Parameters

Not all comparison tasks are equal, and the exact nature of the task will affect the time and cognitive power required to perform a comparison (Just and Carpenter, 1976). However, several trends are clear with respect to comparison tasks. When comparing a scene to some mental criterion, fixation time may be as little as 130 msec., but normally fixations to compare figures and numbers require approximately 200 msec once the viewer is oriented (Just and Carpenter, 1976). If mental manipulation of the figures is required to perform a comparison the fixation time, or subsequent fixation time can increase to upwards of 500 msec (Williams and Harris, 1985; Just and Carpenter, 1976).

These studies demonstrate that fixations used in simple comparison tasks are comparable in duration to fixations occurring in free scanning tasks (< 300 msec/fixation). Free scanning tasks have very low cognitive demand since they are undirected (i.e. - relying on automatic viewing strategies). The cognitive demands for a simple comparison task should also have low cognitive demand based on similar automatic strategies. This would suggest aviators' cross-checks, although perceptually demanding, are not cognitively demanding if the cross-check relies on automatic scanning and comparison strategies. Hazardous states of attentiveness affecting automaticity

should have a significant affect on instrument cross-checks which rely on automatic scanning and/or comparison strategies. Fixation frequency should decrease and length of fixation should increase. The increase in fixation duration may be due to one or two effects. First, fixation duration increases as more long fixations occur for problem solving. Second, fixation duration increases as fewer short fixations occur due to loss of automaticity.

2.14.2. Mental Arithmetic Effects on Psychophysiological Parameters

Like comparison tasks, mental arithmetic can be automatic or quite involved. Eye movement patterns for higher cognitive processes slow with increased degree of difficulty when reading (Rayner and Morris, 1990) and interpreting numbers (Williams and Harris, 1985). In the former case both the fixation on the difficult reading material and the subsequent fixation showed a increase in time. In the latter case, the increased time was accounted for in fixations after digitally formatted numbers were read. Mental arithmetic should result in fixations longer than those used in the comparison tasks of the vehicle operator's cross-check.

Within the context of this study, mental arithmetic tasks were completed as part of the standard navigation tasks. Navigation tasks are planned distractions which are required for safe conduct of vehicle operation. Studies cited above suggested these algorithmic arithmetic tasks would result in longer average fixation times than those fixations used as part of the vehicle operator's cross-check.

2.14.3. Problem Solving Effects on Psychological Parameters

Although problem solving exercises were not specifically incorporated into this study, the realistic nature of the new tasks performed resulted in *ad hoc* solutions by the subjects. Some of these newly learned procedures were practiced numerous times during the study, and could have reached some level of automaticity. Although these activities did not occur uniformly across the subject population, they do constitute a unique source of variation to consider. To account for variance introduced by unique learning experiences, study design included experience level and familiarity with the environment.

2.14.4. Working State of Attentiveness Effects on Psychophysiological Parameters

Before considering quantitative descriptions of hazardous states of attentiveness, normal working patterns must be identified. When engaged in an efficient instrument cross-check, an aviator would be moderately aroused as predicted by Yerkes and Dodson (1908; Figure 2), and confirmed by Easterbrook (1959).

Cognitive activity for a cross-check task remains stable and relatively low, since the nature of the task remains constant throughout the cross-check task. The task is a series of comparison sub-tasks. Comparison tasks are performed rapidly with low cognitive loading (Sternberg, 1975; Just and Carpenter, 1976).

For an aviator, the number of fixations directed toward areas providing useful information should reach a local maximum when the operator is working most efficiently. This maximum occurs because comparison tasks, which comprise the cross-check, do not require significant cognitive resources (Just and Carpenter, 1976). Decreases in

performance would logically be reflected by decreased fixation frequency due to increased fixation time.

2.15. Hazardous States of Attentiveness Effect on Psychophysiological Parameters

One goal of this study was to identify attentiveness. Hazardous states are characterized by dangerous performance. Dangerous performance in aviation is gross deviation from normal airspeed, altitude or course. These deviations could cause a mid-air collision or CFIT. Lesser performance deviations would trigger action by ATC, since ATC limits are designed to ensure aircraft separation should two aircraft deviate simultaneously. It is the third objective of this study to link performance and psychophysiological parameters. Hypothesis concerning the links follows.

2.15.1. Absorption/Distraction

Whether triggered by an internal or external mechanism, the characteristics of absorption and distraction are the same. In both cases, there should be a reduction of fixations per second when the automatic strategies involved in the cross-check are abandoned. With absorption, the cross-check itself is not completely abandoned. An operator increases the information update rate on some small portion of the cross-check to the detriment of other normal viewpoints in the cross-check (Moray and Rotenberg, 1989). This is especially dangerous because aviators may feel they are maintaining a good instrument cross-check, when they have neglected to attend to other important parts of the cross-check. Aviators may or may not be cognizant of attention loss due to distraction. The situation depends on whether focused or divided attention is used.

Absorption and distraction are most dangerous where cross-checked variables are likely to change rapidly.

For example, absorption commonly occurs when coming upon a speed trap on a busy interstate route with your speed in excess of the posted speed limit. When you notice the state trooper in the median, attention is rapidly concentrated on your speedometer while you decelerate to the legal speed limit. Then attention is shifted to another point of absorption as you look in you rear view mirror to determine if the trooper is pulling out to pursue you. You have completely ignored the most important part of your cross-check, the road in front of you. You may even feel a false sense of security since your fixation points are part of the normal vehicle operator cross-check. Still, you are not observing the road in front of you, where the car you were tailgating has slowed dramatically for the same speed trap.

Absorption was demonstrated in a classic eye movement study conducted by Moray and Rotenberg (1989). Subjects attempted to maintain temperature and fluid flow conditions in a process control model. When a faulty valve on one of four fluid baths reduced fluid flow without any fault indication the subject was expected to detect the problem by observing changes in the fluid level for that particular fluid bath. In a different scenario, a second faulty valve was introduced after the initial fault. A number of interesting results occurred.

1. The amount of time devoted to observing the initial fault area increased three fold when the fault was noticed. Fixation time was shifted from other targets, but the scan pattern still covered those targets.

2. Subjects fixated on both the first fault and second fault within five seconds of the fault occurrence.

3. Subjects responded to the initial fault within 20 seconds (50% response level), whereas it took 40 seconds (50% response level) to respond to the second fault.

Summary - Absorption and distraction occur when aviators are highly aroused. Fixation times increase because cognitive activity is high. Study results indicate subjects perceive but do not respond to performance errors. The portion of the HIP Model affected by absorption is not perception, but decision and response selection.

> *Hypotheses - The hazardous states of attentiveness, absorption and distraction, occurring when the aviator is highly aroused, will result in an elevated Index of Engagement, and will be characterized by a significant change in the eye gaze dwell time (>5%) (Moray and Rotenberg, 1989).*

2.15.2. Vigilance Decrement

Vigilance decrement is caused by lack of sensory stimulation. Airline operations over long periods of time normally reach some steady state over the course of the journey. Studies have demonstrated that vigilance decrement is not a concern for the first hour of operation (Akerstedt, Torsvall, and Gillberg, 1987). After the first hour, if changes in visual stimuli slow to match the normally lethargic rate of change in auditory and tactile stimuli, boredom becomes a factor.

Cases like this are common in airline operation. For example, in instrument meteorological conditions a pilot sees only gray fog when looking out the cockpit windows. When the autopilot is engaged airspeed, altitude and course are stable, the only

display changing is the mileage countdown to the next waypoint. This display change occurs at a constant, slow rate.

If the primary cross-check is boring, the aviator will slow the cross-check update rate and seek other visual stimulation, just as the captain of Cali, Columbia crash demonstrated. The vehicle operator searches for stimuli beyond the required instrument displays. Therefore, fixations no longer follow a normal working transition pattern but become increasingly random as in free scanning. Fixation frequency may be similar to that of free scanning, and/or instrument cross-check since the operator would be in a free search mode. However, fixation frequency may slow slightly since the free scanning behavior would result in significantly longer saccades to move vision away from primary instrument displays. As a vigilance decrement occurs, fixation frequency should drop.

Summary - The human visual system naturally seeks out new, and unique information (Biederman et al, 1981). Although there are no studies documenting changes in scan patterns caused by boredom, it is reasonable to assume the "bored" scan pattern would depart the area(s) of interest in an effort to seek out new and unique information. Lacking an underlying scan strategy, the resulting scan pattern would be a stratified random scan pattern, dwelling on areas of high visual interest with greater probability (Harris, 1990; Cole and Hughes, 1990).

> *Hypothesis - The hazardous state of attentiveness, vigilance decrement, occurring when the aviator is in a low state of arousal, will be characterized by cognitive activity similar to that of a normal cross- check, and will be characterized by a significant change in scan strategy on the primary instrument displays (>50% reduction; Moray and Rotenberg, 1989).*

CHAPTER 3.

METHOD

This study was conducted at NASA Langley Research Center with the support of the Crew Vehicle Interface Group. The Advanced Civil Transport Simulator (ACTS) served as the platform for the study. This simulator was created, in part, to act as a test bed for advanced flight deck concepts which are now employed in commercial production aircraft (McDonnell-Douglas MD - 11 and the Boeing 777). Subjects felt the ACTS provided a realistic work environment, and were highly motivated to participate in the study since they were able to test their piloting skills on state of the art operational equipment. Greater detail on the ACTS, and other hardware used in this study, is available through NASA Langley.

The purpose of this study was to place subjects in an operational setting and record psychophysiological measurements at different levels of arousal and states of attention while subjects performed realistic aviation tasks. The scenario selected was an instrument proficiency training sortie which allowed for manipulation of workload to create high, nominal, and monitoring workloads, described below. Subjects flew in the left (captains) seat, while the copilot flew in the right seat. Workload was changed by the copilot in accordance with the script in Appendix C.

Workload was a subjective measure of the subjects' ability to complete the tasks assigned in the script. Monitoring workload was a supervision of the autopilot or copilot. The Nominal workload was a basic instrument flying task flown by the subject. The High Load workload was a challenging flying task flown by the subject. High Load workload

tasks were developed and tested by a group of three experienced aviators to ensure the tasks would challenge even the flight rated subjects.

3.1. Tasks

During the prebrief subjects were informed their primary mission was to maintain the aircraft simulation on the proper airspeed, on the proper altitude, and on course. The primary means of accomplishing this task in Instrument Meteorological Conditions (IMC) is by cross checking the flight deck instruments which provide aircraft performance data. One secondary task, completion of computation worksheets, was a common distraction for all subjects. One computation worksheet was completed during each six minute simulation segment. Other tertiary tasks were added to the scenario after familiarization to achieve the desired workload. These tasks were not changed during the simulation, making the workload increase constant across all segments. Tasks are described later under Workload Manipulations.

3.1.1. Aircraft Simulator Cross-check Requirements

All subjects monitored, or performed a limited number of routine aviation activities centered around an instrument cross-check. Acting as the pilot in command, each subject's primary goal was to maintain their simulation on the desired airspeed, altitude and course. The desired aircraft parameters were available on flight deck displays, from the copilot, and from the Air Traffic Control (ATC) controller. As would be the case in a real aviation environment, subjects were reminded of normal operating

parameters, Table 3.1, whenever they exceeded predetermined limits on airspeed, altitude, and course guidance.

Table 3.1. Aircraft Performance Limits (Federal Aviation Regulations, 1994)

Phase of Flight/Parameter	**Limits**
Cruise/Altitude	±300 Feet
Cruise/Airspeed	±20 Knots
Cruise/Course Deviation	±1 Nautical Mile
Approach/Glidepath Deviation (ILS)	±2 °, or 50 feet for non-precision approaches
Approach/Course Deviation (Localizer)	±2°
Approach/Airspeed	±10 Knots*

*Arbitrary - appropriate limits to approach airspeed deviation vary by aircraft type

Performance parameters, altitude error, airspeed error, and course error, were recorded at 1 Hz. The goal of this study was to determine psychophysiological measures to model variety of operational performance levels. The first objective to accomplishing that goal was to provide subjective workload conditions that would result in a variety of objective performance levels. The performance levels were not strictly controlled since the study was to measure psychophysiological parameters under operational conditions. However, some controls were exercised to ensure all subjects stayed within normal operating conditions. Specific controls are detailed later.

3.1.2. Computational Tasks

In addition to performing an instrument cross-check, all subjects completed computational exercises related to normal aviation. The computational exercises provided a second locus of eye fixations outside of the primary instrument displays. In addition, the computational exercises provided a distraction requiring intense cognitive exercise in which automaticity could not be employed. Although familiar with the concepts these exercises were based on, none of the subjects had performed like calculations in an aviation environment. Most subjects complained, saying, "That's what they make calculators for!"

3.2. Subjects

Subject ages ranged from 27 to 44. Four subject pools were used, each group contained one female subject and three male subjects. Due to the variety of skill levels and levels of familiarity in general aviaiton, a 2 x 2 experimental design was utilized (Table 3.2). Two skill levels, rated pilot versus unrated, and two currency levels commercial transport qualified/familiar, and non-transport qualified/familiar. Four subjects were drawn from each pool for a total of 16 subjects. Familiarity with the environment was treated because of its potential to affect subjects' level of arousal. All subjects were four year college graduates who reported normal, or corrected to normal vision (Table 3.3). Visual acuity was verified using displays of symbols of known visual angle.

Table 3.2. 2 x 2 Design of Experiment (Skill Level x Familiarity with Environment)

	Transport Familiar	**Not Transport Familiar**
Pilot	PF - Airline Pilots	PN - Airforce Pilots
Unrated Aviator	UF - NASA Employees	UN - Public at Large

All subjects were volunteers. The participation of male subjects depended on the individual's availability relative to the simulator schedule. However, the female population available for the study was limited. Female subjects were matched to the first available simulation period. Subject profiles are contained in Table 3.3. Gender was omitted from profiles since there was only one female per group. Identifying subjects by group and gender would allow identification of individual subjects.

The average age for subject groups was not significantly different. Age ranged from 31 for the NASA technicians, to 40 for the commercial airline pilots. Variance in age was greater for the novice group ($p<0.05$). This group was drawn from the general public having contact with personnel working at NASA Langley.

The pilot qualified, transport familiar (PF) subjects were provided by contract to NASA. The PF subjects were qualified in different type aircraft (B-727/B-757/B-777/MD-11), and had flown in the previous week. The pilot qualified, non-transport familiar (PN) subjects were drawn from a pool of approximately 100 US Air Force aviators who had no commercial aviation experience. The Air Force aviators had flown various aircraft including the T-37, T-38, A-10, F-4, F-15, F-16, C-23, and KC-135. All Air Force aircraft flown, with the exception of the KC-135, are much smaller, and

possess different flying characteristics form the commercial transports. The KC-135, used in aerial refueling, is a heavily modified Boeing 707 without stability augmentation, making its handling characteristics and instrumentation significantly different from those of the above commercial transport aircraft simulator.

Table 3.3. Subject Profiles

Subject #	Age	Commercial Aviation Exp	Aviation Hours	Simulation Hours	Reported Visual Acuity
1	44	No	N\A	50	20/20
2	27	No	N\A	40	20/20
3	39	No	N\A	0	20/15
4	27	No	N\A	2	20/80 c 20/20
5	27	Yes	N\A	30	20/20
6	29	Yes	N\A	112	20/30 c 20/20
7	39	Yes	N\A	400	20/20
8	29	Yes	N\A	60	20/40 c 20/20
9	42	No	3,000	800	20/25 c 20/20
10	43	No	3,300	200	20/15
11	37	No	2,100	340	20/20
12	34	No	2,600	402	20/15
13	44	Yes	5,000	10,000	20/20
14	40	Yes	12,000	340	20/20
15	37	Yes	4,200	200	20/20
16	39	Yes	8,700	200	20/40 c 20/20

c – vision corrected

Unrated, commercial transport familiar (UF) subjects were drawn from a pool of approximately 30 NASA employees who had worked and flown in the transport simulators at NASA Langley. All had been in the ACTS simulator while it was operating. Unrated, non-transport familiar subjects (UN) were drawn from the public, at large, and from a pool of NASA employees with no simulator familiarity. Some of the

unrated subjects had general aviation experience but none had ever possessed any type of commercial aviation rating.

3.3. Apparatus

Hardware used to conduct the study included the ACTS simulator, the Crew Response and Evaluation Window (CREW), a Cadwell Brainmapping ensemble with remote control, an ASL oculometer with remote control, and a remote Air Traffic Control (ATC) station. The experimenter was in contact with all remote sites via headset on a discrete communications circuit. A wiring diagram may be found in Figure 3.1.

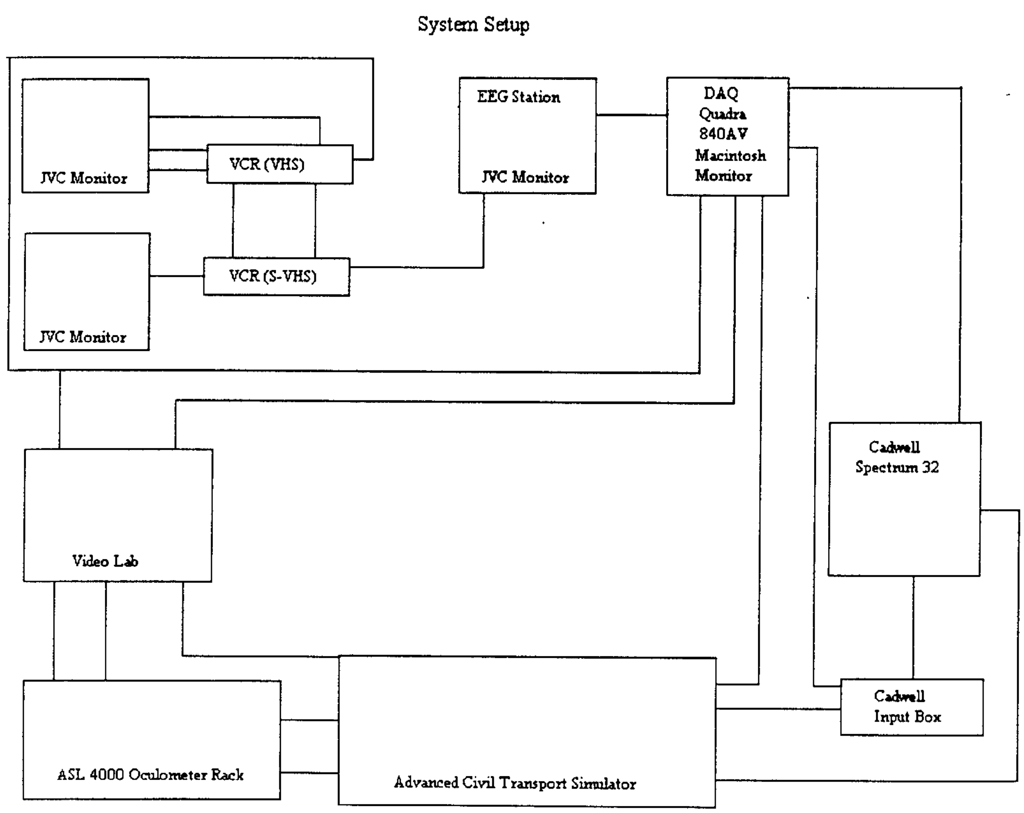

Figure 3.1 ACTS/Instrument System Set-up

3.3.1. The Advanced Civil Transport Simulator (ACTS)

The ACTS is a two place (pilot/copilot) flight deck with a forward looking out the window graphical interface provided by a Silicon Graphics Onyx. Graphics resolution was sufficient to allow taxiing around an aerodrome, takeoff, approach and landing; simulation of instrument meteorological conditions was very convincing. Flight deck accommodations (Figure 3.2) were similar to those on a modern (MD-11/B-777) operational flight deck down to the aircrew seating, with two exceptions. First, instead of the control yoke normally used to control inflight attitude of American made commercial aircraft, this simulator used an advanced concept side stick controller. Second, the flight engineer position, behind the copilot, was occupied by a silicone graphics terminal used to control three computer routines necessary for the simulation. This station was occupied by a NASA technician for all simulator sessions. On several occasions this technician was called upon to create a more difficult simulation platform for aviators who were not challenged by the normal simulation profile.

Two Unix based routines controlled out the window graphics, and aerodynamic modeling respectively. The third routine, a PC based routine which controlled the flight deck displays, was coordinated with the other two to create the overall simulation. Real time control from the flight engineer station allowed changes in the weather via out the window graphics and through the aerodynamic model if winds or turbulence were desired.

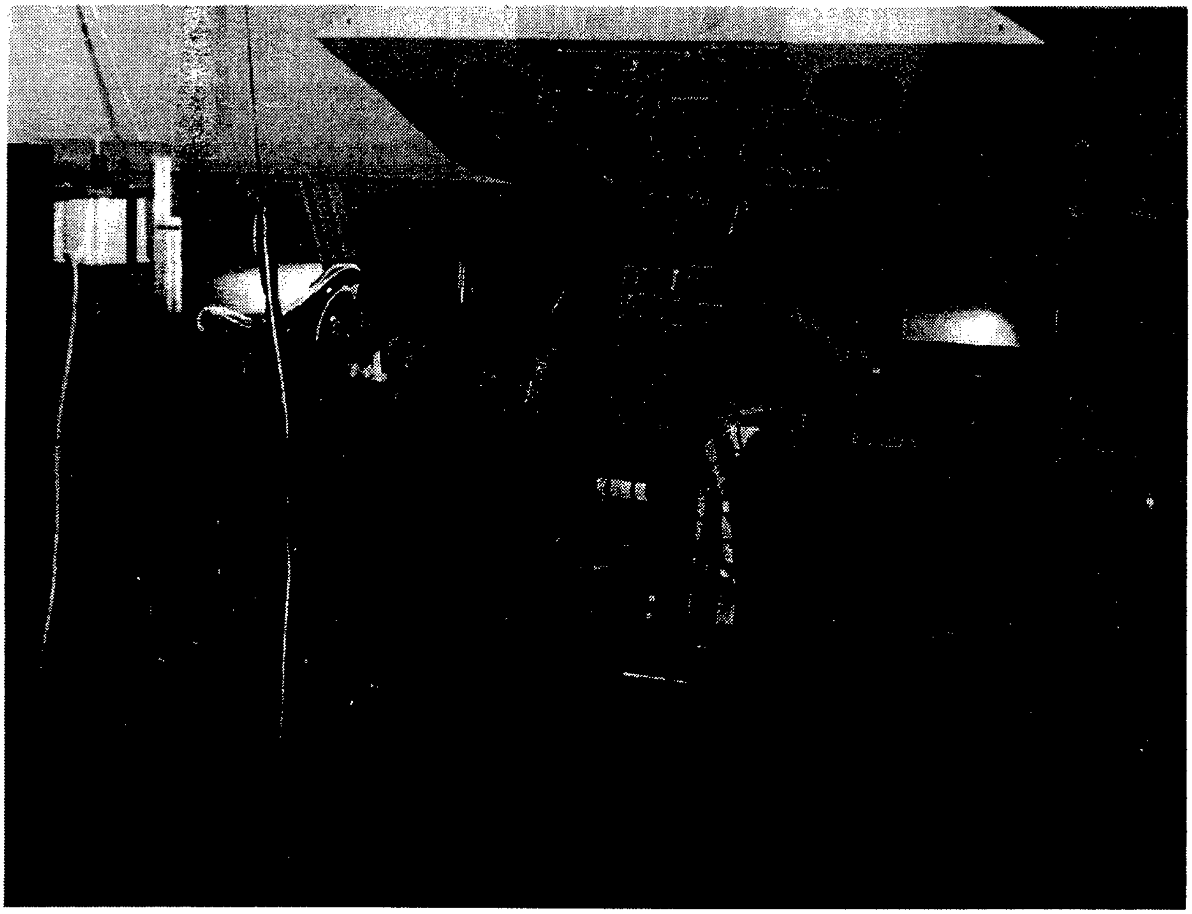

Figure 3.2. Flight Deck Accommodations

3.3.2. Aircrew Display Overview

The display console in the ACTS had five displays. Directly in front of each pilot was the primary instrument display described below. The instrument displays were furthest outboard. Inboard of the instrument displays were the system displays. Other information required by the pilot but not part of the normal instrument cross-check could be displayed on these reconfigurable system displays. In the center of the console was the Airborne Caution And Warning System (ACAWS) display. The ACAWS display, and

all other panels seen in Figure 3.1, were operational for the simulation, but were not necessary for the subjects to complete this study.

Displays developed for the ACTS incorporated numerous advanced display concepts. Generally, primary performance information (airspeed, altitude, and heading) was displayed in digital format. Airspeed and altitude were displayed on tapes seen to the sides of the primary instrument display (Figure 3.3). Course information for all phases of flight was displayed at the center of the display. The actual displacement from the course (cross track error) was displayed in digital format in the lower half of the display on the Horizontal Situation Indicator (HSI).

Trend information was displayed in analog form immediately adjacent to related performance information. For example, trend bars for altitude and airspeed were located outside of the performance information to allow for rapid crosschecking of the information. The zero points for both trend indicators were adjcent to the center display box for their respective performance indicators with increasing trends showing in the upward direction, and decreasing trends showing in the downward direction. The larger the trend, the longer the analog bar.

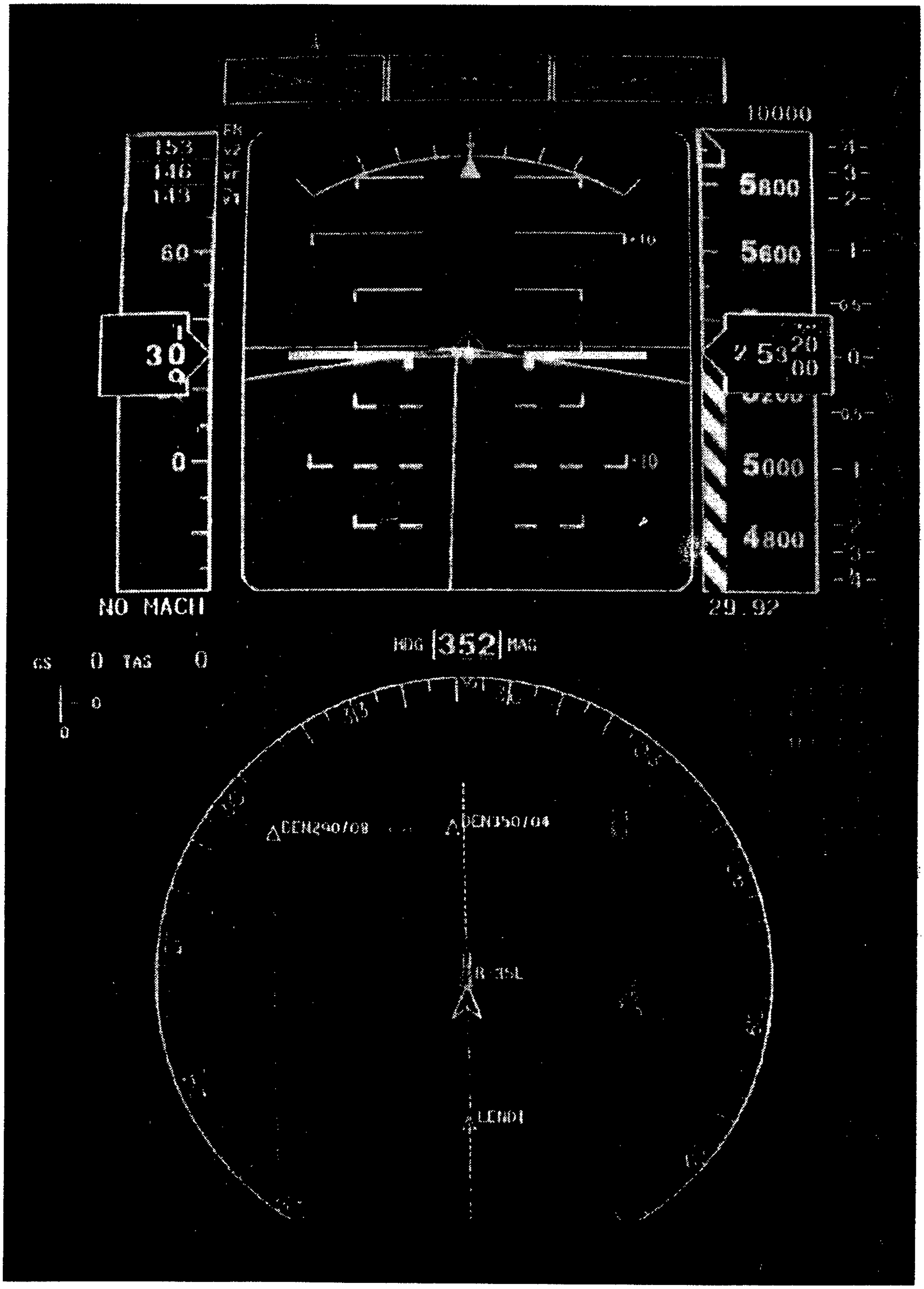

Figure 3.3. Primary Instrument Display

The trend information for course maintenance was immediately below the course marker at the bottom of the attitude indicator (top half of Figure 3.3). As suggested by the results of other studies, the course and trend information were displayed in an external reference format (Fitts and Jones, 1950; Wickens, 1992). For example, if the aircraft had been inadvertantly turned 10° to the the right of the heading necessary to maintain course, the course indicator would begin to drift to the left, while the course trend analog bar would display a value to the right of the course indicator. If the aircraft was then turned 10° left to parallel course, the course indicator would remain steady but displaced to the left. Another 10° left turn would result in the course indicator drifting back toward the center of the display and the trend indicator analog bar would be displayed to the left, toward the course line.

The Horizontal Situation Indicator (HSI) filled the lower portion of the primary instrument display (Figure 3.3). The forward point of the delta winged aircraft figure represented the aircraft simulator's actual position; in this case the aircraft was at the end of the runway ready for takeoff. The white dotted line extending from the front of the aircraft symbol was a trend vector. If the aircraft was in a left bank, the trend vector would curve left to project the turn.

The information block in the upper right corner of the HSI shows the primary instrument parameters in digital format. Triangles displayed in white indicate points on the planned route of flight as entered into the aircraft simulator navigation system. The light grey lines forming the rectangular circuit represent the desired course line segments between the course points. Turns had to be initiated prior to the course points to maintain the desired course line. The planned turns assumed 20° of bank at an airspeed of 220

Knots and an altitude of 10,000 feet. The circles indicated the geographic position of significant points in the vertical flight profile (TOC-Top Of Climb, TOD-Top of Descent, and BOD-Bottom of Descent).

Heading was displayed in two formats on the HSI. The traditional compass rose was displayed with the heading marker at the top of the rose, and a digital repeater of the magnetic heading provided redundancy above the compass rose. Finally, winds affecting the aircraft simulation were broken into along track and cross track components, and these components were displayed in the upper left corner of the HSI.

The view shown in Figure 3.4 was the copilots view of the system and ACAWS displays to the left of the instrument display. The system display contains engine information on the top half of the right display and had a field for optional system information on the bottom half of the display. The example below shows the copilot's display with the fuel system graphic. Graphics of the electrical systems, air circulation system and hydraulic systems were also available. Engine intruments were displayed in pairs. Typically, the subject looked only at the pair of dials in the upper left corner of the display. These dials displayed Engine Pressure Ratio (EPR), the thrust generated by the aircraft simulator's engines.

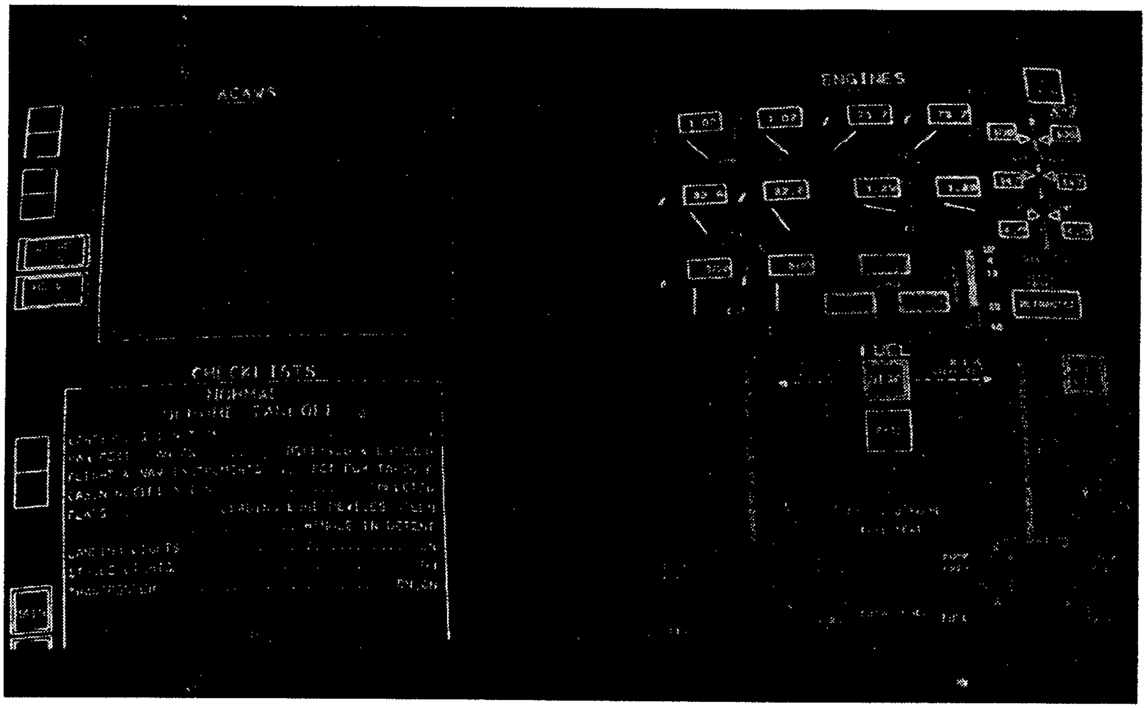

Figure 3.4. ACAWS and System Displays

The ACAWS display (Top half of left display, Figure 3.4) was part of the center display used by both pilots. The top half of the display showed cautions and warning if any system anomalies were encountered. The bottom half allowed display of any normal, or emergency checklists. Neither the ACAWS, nor the checklists were necessary for the subjects to complete this study, but they did act as distracters. The copilot was responsible for all anomalies and checklists.

3.3.3. Crew Response Evaluation Window (CREW)

A Macintosh Quadra computer using NASA Langley developed software (Labview Code), accepted two video inputs, two audio inputs, and six discrete inputs to create an integrated display. In this study pupil video was superimposed on gaze point

video from the oculometer and fed to the CREW as the first video input (right monitor Figure 3.6). The second CREW video feed was an over the shoulder camera focused on the captain's primary instrument displays (left monitor Figure 3.6). All video input was time stamped before reaching the CREW display. Audio input from the experimenter's discrete audio circuit, and an area microphone inside the simulator cab comprised the two audio sources. The six discrete signals fed to the CREW through the Cadwell were:

1 - 3. Three tone indicators for evoked response (not used in this analysis),

4. An event trigger which could be set by the experimentor,

5. A segment counter which indexed progress through the simulation profile, and

6. A wind/turbulance indicator.

Finally, a separate discrete line for peripheral temperature data went directly to the CREW from the subject position in the ACTS. The system setup diagram, Figure 3.3, displays the CREW setup positioned in the top center portion of the figure. A sample of the CREW display may be found in Figure 3.5

Index of Engagement and Peripheral Temperature Data Records. The CREW computed an instantaneous Index of Engagement every second by averaging the one hertz index samples over the previous 20 seconds. Data was recorded digitally and a 12 minute strip chart readout also displayed this value every four seconds to allow detection of trends during the simulation. A second twelve minute strip chart displayed peripheral skin temperature measured by a thermistor taped on the center of the dorsal surface on the proximal portion of the subjects left index finger. Peripheral Temperature data were also digitally recorded at two hertz. Twelve minute strip charts corresponded

to the time required to complete one circuit arount the instrument pattern used for this simulation.

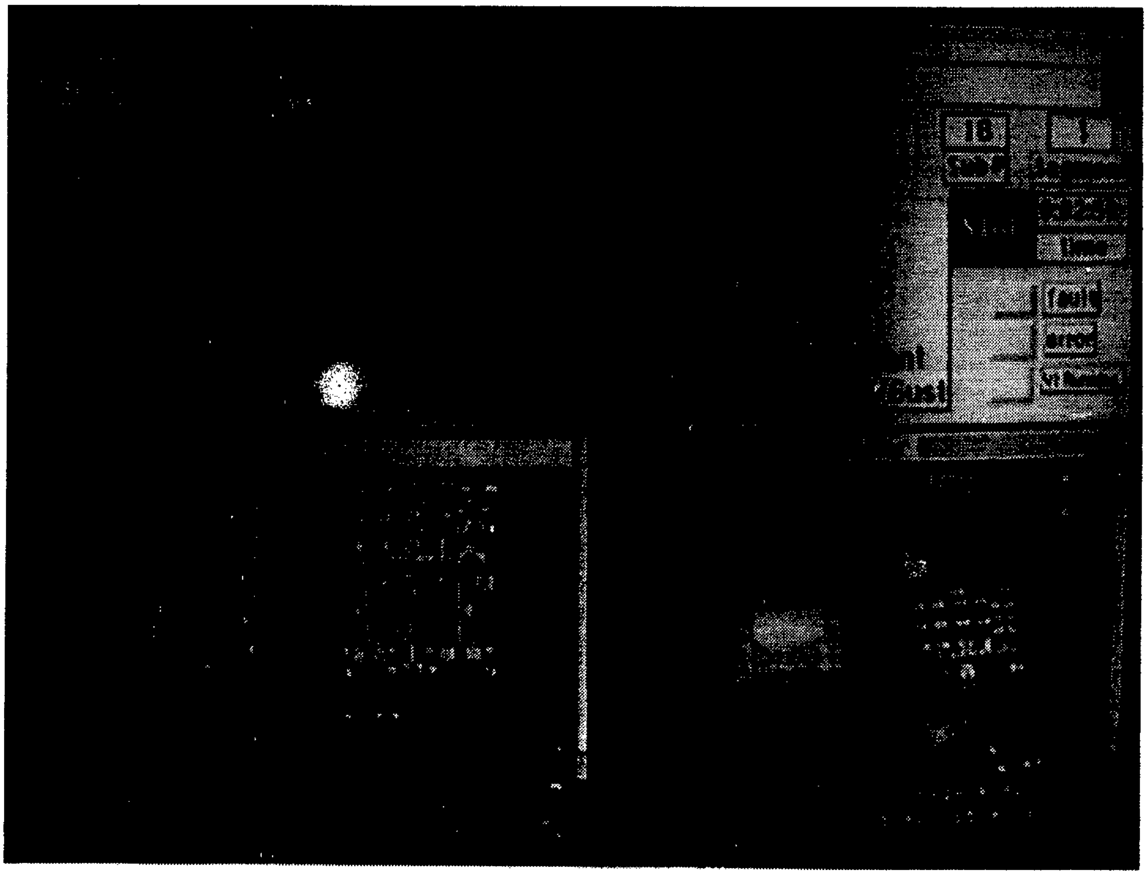

Figure 3.5. Crew Response Evaluation Window/CREW Display

The CREW display composite video was recorded in an S-VHS format, and the video output of the oculometer gaze point was recorded in VHS format. A video feed was taken from the S-VHS recorder to provide a real time display of physiological data to the experimenter. The experimenter's display was at floor level to the right of his seat. The experimentor's seat occluded the subject's line of sight to the display.

3.3.4. Cadwell Spectrum 32

The Cadwell brainmapper (Cadwell Laboratories, Kennewick, WA) has the ability to process a full EEG ensemble. However, a reduced number of electrical sites (11) was used on all subjects to accommodate placement of a head mounted oculometer. All sites available were prepared, and recorded at 2000 Hz. However, only the central parietal site (PZ) was passed to the CREW to be used for calculation of Index of Engagement. The Cadwell also recorded the six discrete outputs mentioned above.

In addition to recording the raw EEG data, the Cadwell processed the raw data through active band pass filters and decomposed it into frequency components by power spectral analysis using fast fourier transforms. The applicable frequency band values were then exported to the CREW for further processing into Index of Engagement. Finally, raw data was saved to the optical disk.

The Cadwell was located adjacent to the ACTS where a technician remained on headset during testing. The EEG monitoring station was connected to the head box via under floor cabling, and the head box was secured on a special mounting bracket affixed to the aft, lower, left portion of the subject's seat. This position was necessary to provide shielding from electromagnetic interference caused by a magnetic head tracker, while maintaining close proximity for electrical hookup.

3.3.5. Oculometer

An Applied Sciences Laboratory (ASL) Oculometer, Model 4250D, was used to record eye movement and pupil diameter from the left eye. The oculometer had been modified from the original ASL configuration to provide better balance on the head.

In the head mounted configuration, an infrared illuminator was shined down from its mount on the headband and reflected off of a beam splitter positioned at approximately a 45° angle in front of the subject's left eye. As the infrared light passed through the various surfaces of the eye some energy was reflected back to be collected by a CCD camera looking through the same beam splitter. The largest reflection was off of the anterior surface of the cornea subtended by the pupil. The brightest reflection resulted from the the point at which light was focused on the posterior surface of the lens. Before operation of the oculometer, a calibration was performed by asking the subject to observe nine known positions while the geometry of the two previously mentioned reflections was recorded. A regression of the nine points was performed to produce equations converting reflection geometry into a look point vector in the coordinate system in which the calibration targets were defined.

Two geometric transformations were necessary to obtain the subject's look point. First, a magnetic head tracker measured three angular rotations, and a point in space relative to the origin of the head tracking system. After the head tracker determined the subject's head/eye position and orientation vectors, the final look point vector was determined by adding the eye tracker vector onto the head tracker position and vector. This produced a gaze vector which intersected a surface at some gaze point.

The second geometric transformation involved defining the surfaces upon which the gaze vector fell. The geometry of the surface had to be defined from the origin of the head tracking system to determine the exact point where the gaze vector and surface intersected. To accomplish this measurement, the oculometer system employed a laser

wand mounted at the origin of the head tracking system. The wand was used to measure the distance and angles to three points on each viewing plane to define that plane.

3.3.6. ATC Control Center

The Langley Mission-Oriented-Terminal-Area Simulation (MOTAS) Facility provided the hardware necessary for monitoring, communication, and direction of the simulation from an Air Traffic Control (ATC) standpoint. The controller station was centered around a Silicon Graphics display which was custom designed to allow the controller to monitor an instrument proficiency flight profile, provide in flight vectoring, and direct a Precision Approach from Radar (PAR). The station also contained multiple communication circuits and a voice disguiser to allow simulation of multiple ATC functions by one controller.

MOTAS was located in a separate, secure portion of the simulation building. The controller used three physically separated speakers to aid in identification of the communication source. One of the speakers was part of the discrete communications circuit employed by the experimentor. The experimentor could initiate changes and corrections to the flight profile on a real time basis by communicating discretely with the controller. Further information on MOTAS is available in NASA technical papers (Credeur et al, 1993).

3.4. Procedure

Each subject completed two simulation study periods on the same day. To minimize variation in subjects' performance levels due to of time of day, all subjects were

run on the same planned schedule. Some subjects started their initial simulation up to one hour late, due to technical difficulties in simulator start-up, subject briefing, or physiological measurement preparation. Afternoon simulation start times were more consistent due to the time buffer provided by lunch. The schedule was:

0730 - 0800	Prebriefing
0805 - 0820	Simulator Familiarization
0825 - 0855	EEG Preparation
0900 - 0930	Oculometer Preparation/Calibration
0930 - 1115	Simulator Period 1
1130 - 1230	Lunch
1230 - 1300	EEG/Oculometer Preparation/Calibration
1300 - 1500	Simulator Period 2
1500 - 1530	Subject Cleanup/Technical Debrief
1530 - 1600	Subject Debrief

3.4.1. Prebriefing

The briefing room was a small conference room with a white board on which the instrument flight pattern and administrative information were presented (Figure 3.6). After the daily schedule was reviewed, subjects were briefed on the equipment to be used, its operational characteristics, and any techniques that might help maintain their comfort without affecting the study. The type of data and means by which it was to be collected, were explained.

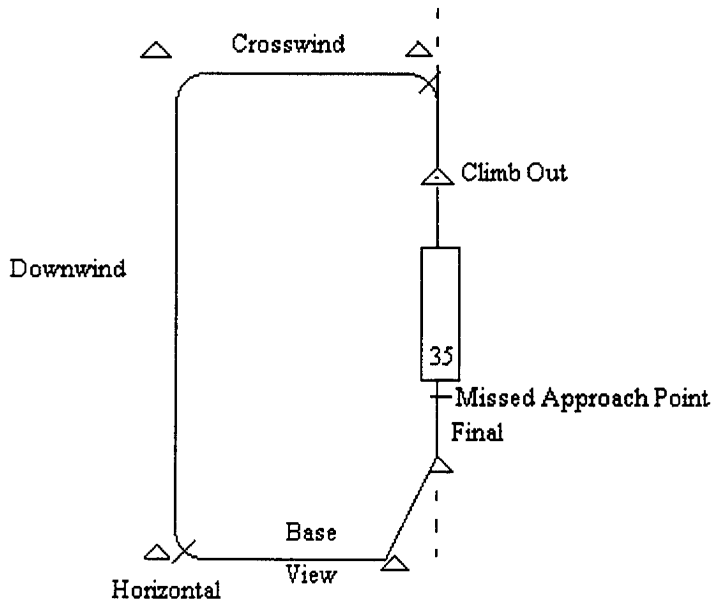

Figure 3.6. Horizontal View of the Instrument Pattern

Next, the experimenter ensured subjects understood the purpose of the study, which was to establish an analytical relationship between psychophysiological measures and aviation performance. First, different states of attentiveness (working, distracted, absorbed, bored, sleepy) were described. Examples were given in general terms such that the expected physiological manifestations of these states were not described. Finally, subjects were asked not to attempt to skew the study by trying to maintain an unnecessarily high level of alertness or attentiveness. It was explained that samples of both attentiveness and inattentiveness were necessary for completion of the study.

The subjects, acting as pilot-in-command, were briefed to assume their primary duty was maintaining the aircraft simulator airspeed, altitude, and course. They were informed that the copilot perform all other crew duties necessary to complete the simulation scenario. In some situations, when the subject demonstrated proficiency in

assigned tasks, additional tasks would be transferred to the subject. It was understood this was not the method for division of labor in commercial aviation.

Since the skill level of subjects was expected to differ, subjects were briefed to expect changes in workload after the simulator familiarization. They were assured the workload manipulations were not a reflection of their performance. In fact, it was stressed that in high workload segments some deviation from nominal performance was expected and required to ensure the subject was task saturated. Although their ability to maintain the simulator on airspeed, altitude and heading was being measured, they should not be concerned if their performance was not perfect.

Before departing for the simulator orientation, subjects were briefed on safety issues related to three aspects of the study:

(1) EEG electrode placement and the potential for skin irritation,

(2) Infrared illumination safety limits for the eye, and

(3) Hot spots caused by placement of the oculometer over the EEG cap.

These briefing items were required as outlined in the Institutional Review Board's approval letter (Appendix A). Subjects also completed an Informed Consent form (Appendix A) as part of this process.

3.4.2. Simulator Familiarization

Before encumbering subjects with EEG and oculometer equipment, they were familiarized with the simulator environment. This simulator checkout served four purposes. First, it allowed the subject to gradually build up familiarity with the new environment in an effort to minimize the stress associated with new experiences. Second,

it allowed the subject to view the flight deck environment without the encumbrance of EEG equipment, and the oculometer. When wearing such equipment subjects tend to move less naturally for fear of damaging the equipment. This restricts their ability to explore the physical layout of the environment, which compromises safety in the event of an emergency evacuation. Third, it allowed the experimenter to ensure all aspects of the simulator profile operated properly prior to beginning the simulator period. Finally, subjects' self reported visual acuity was verified using simulator displays. Subjects were required to read numbers on the Vertical Velocity Indicator (VVI). The VVI had the smallest character size on the primary instrument display, with characters subtending 0.5° visual angle. Then subjects were asked to read smaller characters in the systems fuel display equating to five minutes of arc with detail subtending one minute of arc, or 20/20 vision. Once the experimenter verified subjects' visual acuity, the preparation process continued.

3.4.3. EEG Preparation

After completing the simulator familiarization, subjects proceeded to the Human Engineering Methods (HEM) Lab located adjacent to the simulator complex. A NASA technician placed a specially modified skull cap with 11 electrode sites on the subjects, and applied electrode gel as necessary to reduce impedance below 5 mohms. A ground site on the cap and two earlobe reference sites were also prepared to the above impedance criteria.

While the technician prepared the subject, the experimenter and subject reviewed a practice set of computation sheets (Appendix B). These sheets were consistent in

format with those used in the simulator. The starting figures and answers for each type of sheet contained the same number of characters to minimize variation in reading patterns. The subject completed one example of each of the five types of computation sheet. The subjects were encouraged to ask any questions concerning the computational tasks, and the sheets were corrected to 100% accuracy. This procedure took approximately 30 minutes.

3.4.4. Oculometer Preparation/Calibration

With the skull cap prepared, the subject returned to the ACTS for placement of the head mounted oculometer. A NASA technician placed the oculometer on the subject's and adjusted the infrared illuminator and cameras to positions providing the brightest pupil reflection. Since the oculometer control panel was remotely located, a second technician maintained radio contact via FM headset with the technician in the ACTS. Two displays were used in the alignment process. The first was a video feed from the pupil camera. This display was used to center the pupil in the camera's field of view. The second was a video feed of the scene camera which was used to center the scene camera on the field of view.

After cameras had been aligned properly, the subject was asked to fixate sequentially on nine static points marked on a cover placed over the primary instrument display. The technician at the oculometer remote control saved the calibration points, while the ACTS technician ensured the subject maintained the conditions best for the calibration process. The subjects were required to maintain a static head position and gaze point, while the points were recorded. A second sweep of the calibration points was

performed to ensure the initial calibration was true. Adjustments were made as necessary to ensure the initial calibration provided a gaze point within 0.5° of the desired target.

At the completion of the calibration process, the experimenter checked remote data displays for the EEG and oculometer to ensure time hacks were synchronized with the master computer running the simulation. After the experimenter was seated in the copilot seat, the ATC circuit and all stations on the discrete communication circuit were checked to ensure clear communications. Finally, each station (Cadwell/CREW, Oculometer, and MOTAS) was polled to confirm they were prepared for the start of the simulation. A one minute countdown was initiated to allow for proper initiation of the CREW display. The final ten seconds were counted down on the discrete communication circuit and the simulation was started at zero hours elapsed time.

3.4.5. Morning Simulation

The first Simulation Session was conducted when subjects should have been at a peak alertness period (Astrand and Rodahl, 1986). In addition, subjects were very aroused by the unique opportunity to fly the state-of-the-art simulator equipment.

The script for Morning Session (SS-1) is found in Appendix D. Both simulation studies were divided into segments related to specific task types. For example, during six minutes of an instrument pattern, constant altitude and airspeed were maintained while the crosswind and downwind portions of the instrument pattern were changed. The remaining six minutes of the instrument pattern were composed of the base, final, and climbout portions of the instrument pattern, in which altitude and airspeed were

constantly changing. Different cross-checks and update rates were necessary for these differing tasks.

Simulation Tasks. The morning simulation consisted of 19 segments made up of seven distinct tasks. Every type of task was completed by the subject at least once in the first hour. For realism, the simulation was started on the ground at the airport gate. The first three segments were ground segments for engine start, taxi, and takeoff. The workload in these segments started at a low level and became progressively more difficult, culminating with a takeoff into difficult weather conditions. The remaining four segment types were airborne instrument flight segments consisting of a cruise condition (crosswind/downwind) and three different types of instrument approaches (base/final).

Engine start. The engine start was accomplished by the copilot. The pilot was required to monitor the engine instruments to ensure operating parameters did not exceed those marked on the display dials. As part of this segment, the subject made a radio call to Denver Ground Control requesting permission for engine start. This allowed subjects to become familiar with the radio transmission procedures used throughout the simulation.

Taxi Segment. The taxi segment was initiated with a radio call to Denver Ground requesting clearance to taxi. After receiving clearance, the subject was required to maneuver the aircraft on the Denver Stapleton tarmac using thrust from the aircraft simulator engines and a combination of tiller and rudder pedals for directional control. The tiller is similar to the top of a steering wheel; it is used to turn the aircraft nosegear when affecting large turns at low speeds (high gain). The rudder pedals also turn the

nosegear when there is weight on the wheels, but these turns are very small (low gain) in comparison; rudder pedals are used for high speed taxiing, takeoff, and landing.

Takeoff Segment. The takeoff segment was initiated with a request for takeoff clearance from the pilot to Denver Tower. When cleared for takeoff the subject (pilot) taxied onto the runway and advanced power to takeoff thrust. This power setting was verified by the experimenter (copilot), allowing the pilot to concentrate on directional control and rotation airspeed. When rotation airspeed was attained the pilot rotated the stick aft to bring the nose of the aircraft up to the commanded takeoff attitude. After breaking ground the copilot retracted the gear and flaps, while the pilot transitioned inside the cockpit to the primary instrument display. The simulator entered the weather at an altitude of 200 feet, approximately one minute from break release.

Cruise Conditions. Cruise conditions are those conditions encountered after the aircraft has completed its climbout after takeoff, but before beginning descent for approach, or approach and landing. (The distinction between *approach*, and *approach and landing* will be described later.) At cruise conditions the aircraft is typically maintained at a constant altitude and airspeed, but course is altered to follow the route specified by the aircraft's flight planned route. Aircraft fly along a number of airways as part of their flight planned route just as ground vehicles travel along a number of interstates or roads as part of their route.

Instrument Approaches. Of the three types of instrument approaches performed in this study, the approach most often used by commercial aircraft today is the Instrument Landing System (ILS) approach. On an ILS approach the pilot is provided with glidepath and runway alignment information. Aircraft computers produce this

information by interpreting beacons from two separate navigation systems on the ground near the end of the runway. The glideslope and course information is then presented on the pilot's instrument display. The pilot maneuvers the aircraft to maintain the ILS indicators centered on the desired position.

The second type of approach segment, a localizer approach, is a less precise method of approaching a runway because it provides no glidepath information. The runway alignment information is the same as that provided on an ILS approach, but altitudes are maintained procedurally based on the distance to the end of the runway. A localizer approach is more memory intensive since the pilot must remember intermediate level off altitudes and their corresponding geographic points while flying the approach.

Third, a Precision Approach from Radar (PAR) is a precise approach using a very different method of navigation. In a PAR, a ground based controller uses a radar to determine the position of the aircraft relative to the runway and directs the pilot to the desired glidepath and runway alignment. The controller literally talks the pilot down to the ground with a series of turns and descent rates. This is a departure from the instrument cross-check used by the pilot in an ILS or Localizer approach, since those approaches are self paced. An PAR requires the pilot to adapt their cross-checks to the control style of the ground controller since you must follow and verify the ground controller instructions.

In this Simulation Session only one <u>approach to landing</u> was performed. The remaining 17 <u>approaches</u> flown terminated when the aircraft reached 200 feet altitude without breaking out of the simulated clouds. This was done for two reasons. First, in an instrument proficiency profile the aircraft will normally perform only low approaches to

reduce wear and tear on the landing gear of the aircraft. Second, this study was built around the pilots' instrument cross-check, but if pilots see the ground they will transition off of the aircraft instruments to the graphics cues provided. Eye movement transitions between visual and instrument conditions is beyond the scope of this study.

Simulation Session-1 Profile. In SS-1 the crosswind/downwind segments were designed to maintain a moderate workload for every circuit of the morning simulation. However, the base/final/climbout segment was designed to vary between high and low workload levels. The cyclic nature of the workload design was selected as representative of the repetitiveness of the real world task. The resulting variations in workload for this Simulation Session are shown in Table 3.4. The script for morning simulation may be found in Appendix C.

Table 3.4. Designed Workload Variation by Segment for Simulation Session-1

Workload/Seg	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
High			X				X				X				X				X
Moderate		X		X		X		X		X		X		X		X		X	
Low	X				X				X				X				X		

SS-1 Downwind/Crosswind Segments. The changes in workload occurred naturally with the flying tasks prescribed in the SS-1 script. With the exception of segment 2, all even numbered segments corresponded to the downwind/crosswind portion of the instrument circuit. Flying the crosswind/downwind segments of an instrument pattern is a nominal workload task. However, all subjects were unfamiliar with the instrumentation and side stick controller of the ACTS. This lack of familiarity created a

workload that was higher for some subjects. In cases where the subjects were task saturated by the basic crosswind/downwind workload, some of the workload changes discussed below were employed to reduce workload.

Other experienced pilots found adaptation to the new instrumentation and controls quite easy. To ensure a nominal workload was maintained for these highly adaptable individuals it was necessary to increase workload. Some of the workload changes discussed below were employed to increase workload. Workload changes were employed between data segments. The basic tasks performed during data segments were the same for all subjects.

SS-1 Approaches. All approaches were flown down to the missed approach/decision height altitude of 200 feet, regardless of the type of approach flown. If the subject could not see the runway environment at 200 feet, a missed approach was to be initiated. Simulator technicians ensured subjects did not see the runway environment during the approach. The ground track and the altitude restrictions at the beginning and end of the approaches did not vary. Approaches always started at 10,000 feet and ended at 200 feet. However, the method of guidance, pilot in command, and autopilot channels employed did vary.

Approaches flown in segment numbers 5, 9, 13, and 17 were designed to be low workload approaches. They were flown by either the copilot or coupled to the autopilot. Although the subject monitored the approaches to ensure the aircraft simulator was on glide slope and course, these approaches were all ILS approaches meeting the highest checkride grading criteria. This situation gave the subject a low workload, when monitoring the approaches.

The approaches flown in segments 7, 11, 15, and 19 were designed to be high workload approaches. Segment seven was a non-precision localizer approach on which the controller purposely made the approach difficult. Segments 11 and 15 were both PAR approaches in which the controller was constantly changing the directions to the pilot (over-controlling). The final approach, segment 19, was the first ILS flown by the subject and winds were added to the approach to increase its degree of difficulty.

Workload Changes. During familiarization, the experimenter monitored performance parameters maintained by the pilot. Workload was increased incrementally until the experimenter observed deviations outside the prescribed performance limits. Once minor performance deviations were observed, the workload was reduced to the previous workload level resulting in nominal performance.

The copilot/experimenter was initially responsible for all cockpit tasks, except for the primary task of flying. These tasks included handling radio calls, selection of navigational aids, and operation of the various autopilot channels. When subjects were too proficient and the desired workload level was not maintained, these tasks were transferred to the subject to increase workload. If these tasks did not create sufficient workload, adverse weather conditions such as turbulence and crosswinds were added.

When subjects experienced difficulty performing the basic aircraft control tasks it was necessary to lighten the load. In these cases, the copilot/experimenter flew the simulator back onto conditions before giving the subject control of the aircraft.

Autopilot Channels. Aircraft simulator altitude, airspeed, and course were programmed in the flight computer. Like all commercial airliners today, this simulator possessed the capability to fly independently when programmed. The autopilot had three

channels, airspeed, altitude, and course. All channels could be engaged allowing the aircraft simulator to fly itself around the instrument pattern, or a limited number of channels, like airspeed only, could be engaged.

3.4.6. Lunch

Upon completion of SS-1 the oculometer was removed from the subject, but the EEG skull cap was retained. (Removal of the skullcap would have required cleansing the contact gel from the scalp before the cap could be properly refitted.) The experimenter and subject proceeded to a nearby briefing room for lunch and debriefing of the first simulation.

Since it one objective was to observe the subject in hazardous states of attentiveness, some steps were taken to depress the subjects level of arousal for the afternoon simulation. A large lunch was provided by a nearby cafeteria, and no caffeinated beverages were permitted. The afternoon simulation profile was explained to the subjects prior to proceeding back to the simulator.

3.4.7. Afternoon Session (SS-2)

The second Simulation Session was designed to take advantage of reduced alertness occurring at a dip in the typical subject's circadian rhythm. The script used for SS-2 may be found in Appendix E. Like SS-1, this Simulation Session was divided into segments roughly six minutes in length, but in SS-2 there were 20 segments.

Simulation Tasks. All segments for SS-2 were airborne segments in the same instrument approach circuit described for SS-1 (Figure 3.3). The simulation was two

hours in length, and all tasks accomplished were unique derivation from the morning tasks. This simulation period began with the aircraft located at cruise altitude beginning the crosswind leg of the instrument proficiency circuit. As with the first simulator study the workload started at a low level. This low workload was a result of the copilot flying the aircraft simulator on autopilot.

Simulation Session-2 Profile. To avoid confounding results, the design of the simulation script (Appendix E) was altered to create a constant, moderate workload on the base/final segments of the afternoon instrument circuits. Variation of workload was designed into the crosswind/downwind segments. The designed workload levels for the second simulator session are shown in Table 3.5. A significant break in the normally sinusoidal pattern of workload variation occurs in segment 11. Following the normal pattern segment 11 would have been a high workload segment, and segment 13 would have been low workload. The designed workload for these two segments was reversed.

Table 3.5. Designed Workload Variation by Segment for Simulator Session-2

Workload/Seg	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
High			X				X						X		X				X	
Moderate		X		X		X		X		X		X		X		X		X		X
Low	X				X				X		X						X			

SS-2 Base/Final Segments. An ILS precision approach was flown on all approaches. A moderate workload level was maintained on these even numbered segments (2,4,6,...,20) by allowing the aircraft simulator autopilot, or the copilot to set up the approach. The pilot (subject) only took over the approach after the aircraft simulator

was established on final course and glidepath. Subjects were quite engaged by the challenge of flying the ILS final approach. Despite their interest in the task, some subjects obviously found this portion of the simulation quite easy. When the experimenter felt the subject was not working to a moderate level, some of the workload manipulations previously discussed were employed to increase subject engagement.

During the afternoon pre-briefing subjects were told an anomaly had been observe in the autopilot operation during previous simulations. (This was true.) Furthermore, they were informed an autopilot check would be performed as part of the afternoon simulator profile. Subjects did not need to know the autopilot check procedure, since the experimenter was familiar with the requirements. During the autopilot check manual controls were used only as part of the missed approach sequence, and in this case only by the experimenter. Subjects were encouraged to closely monitor the approach to observe any anomalies. In this manner subjects were moderately engaged but "hands off" the controls. These approaches in segments 8 and 10 combined with the low workload requirements of segments 9 and 11 created an ebb in attention required.

SS-2 Downwind/Crosswind Segments. The downwind/crosswind cruise segments 1, 5, 9, 11, and 17 were designed for low pilot workload. These segments were flown either completely coupled to the autopilot or by the copilot. The subject was still required to monitor the aircraft simulator position. If the experimenter was in control of the aircraft simulator the flight profile remained very predictable to minimize the subjects perceived monitoring requirements.

Workload was increased in segments 3, 7, 13, 15, and 19 with the help of the controller in MOTAS. In segments 3 and 13 the subject was vectored around the pattern

amid reports of other aircraft intruding upon the planned flight path. As soon as the subject stabilized on the previous vector, the controller commanded a turn to another heading. The demand of an instrument cross-check was more intense in this dynamic situation. Flight planned airspeed and altitude remained constant.

The workload in segments 7 and 15 was increased by constant changes in altitude. As soon as the subject attempted to stabilize their VVI to remain on an altitude, a new altitude change was commanded. This maneuvering tended to be slightly more challenging since it is more difficult to maintain constant airspeed in a climb or descent. Subjects were directed to fly the flight planned route on these two segments.

Segment 19 was a hybrid of the two previous segments. Since this segment included constant changes in both altitude and heading its attention requirements may have been slightly greater, but it was also flown after 3.5 hours of practice. Subjects were directed to maintain their normal airspeed on this segment.

3.4.8. Debriefing

While the subject was being disengaged from the physiological monitoring paraphernalia, the experimenter met with four NASA technicians. These technicians operated the oculometer and EEG equipment, and the ACT controller. The experimenter and technicians discussed and documented any issues with the two simulation sessions which had just been completed. The subject was given a short tour and explanation of equipment operation to facilitate understanding of the study.

In the formal debriefing, the subject provided the following personal data for the record.

1. Age.
2. Visual acuity (uncorrected/corrected).
3. Dominant Eye.
4. Aviation Experience.
5. Aviation simulator experience.
6. Desktop simulator experience (Red Baron, Microsoft Flight Simulator, etc.).
7. Date of last flight.

The experimenter and subject viewed portions of the CREW videotapes on "Fast Forward" to demonstrate how physiological measures changed with workload. The subjects were then asked to describe the different states of attentiveness experienced in each simulation session.

They were primed with the following questions.

1. During the first simulation, were you working hard?
2. Were you bored?
3. Were you distracted?
4. Did you find yourself staring at anything on or off the instrument display?
5. Were you tired or fatigued?
6. Did you feel sleepy?
7. Did you stay alert during the entire simulator session?
8. Is there anything else you would like to tell me about your "states of attentiveness" during this simulator session?

These same questions were asked with reference to the second simulator session.

Chapter 4.

DATA PROCESSING

4.1. Video Tape Review

All Crew Response and Evaluation Window (CREW) video tapes were reviewed to determine initial quality of data. During the initial review simulation start time was recorded. There after, elapsed time was used to coordinate the three different data sources (simulator performance data, EEG/Temperature date, and oculometer data). Video review provided the times at which preprogrammed events occurred for each subject. This allowed selection of data from each subject that was comparable across all subjects.

Each segment was planned to be approximately six minutes in length. Time varied for each segment according to how accurately the subject followed the flight planned route. However, the total time for sixteen subjects to complete the simulation did not vary by more than six minutes over the two hour simulation period. Start time and end time for each segment was recorded to ensure data samples did not spill over into adjacent segments.

Within each segment the study was designed to provide two different data samples. One sample occurred with no distractions. In this one minute period there were no radio transmissions or outside communication, only the primary instrument task. The second sample was designed with a paperwork exercise providing a deliberate distraction (secondary task). The secondary task is explained below.

Total time to complete these paperwork exercises was designed to be one minute, but varied from 40 to 240 seconds based on the subjects' abilities. Since the data segments were broken down into 12 second increments for analysis purposes it was necessary to reduce the number of increments from five (60 seconds) to three (36 seconds) to accommodate subjects requiring only 40 seconds to complete the secondary task.

The method of aircraft control also varied throughout the simulations, but was consistent across subjects. As in a normal aviation scenario, portions were flown with the pilot controlling the aircraft manually, the co-pilot controlling the aircraft manually, and with various modes of the autopilot engaged. There was concern that manual co-pilot (experimenter) control could produce results substantially different from use of autopilot. Results show this was not the case.

During the video review, the time of each change of control mode was annotated using the following list of control codes:

SM - Subject flown manually
SAA - Subject flown with autopilot controlled airspeed
SAC - Subject flown with full autopilot control
EM - Experimenter flown manually
EAC - Experimenter flown with full autopilot control.

This method of video analysis allowed selection of data segments that were similar across all subjects.

Video tapes also provided a backup for some forms of data which were recorded during the simulation. Spurious inputs to the temperature data for subjects one through ten required that average peripheral temperatures for all subjects be determined from the CREW video. Video also provided verification of other data sources.

4.2. Factor Data

The experimental factors associated with data were recorded with each data set. These factors included subject number, simulation segment, subject gender, subject experience in the commercial aviation environment, subject aviation rating, workload rating (level of difficulty), and time of day.

4.3. Performance Data

Data was processed using code written in Quick Basic. Performance data was recorded from the simulator at a rate of eight samples per second. Other data rates were significantly higher therefore it was not possible to display all data samples and simultaneously examine data trends over the two hour period. For ease of analysis data was averaged over twelve second intervals which allowed display of an entire two hour simulator period on a computer screen.

To determine trends in performance error, each axis of control error (airspeed, altitude, and course) was continuously plotted for each subject over the entire simulation period. This also allowed detection of anomalous data. In addition, the absolute magnitude of stick and throttle control inputs was plotted for the three control axes.

The flight planned route for the entire simulation period was preplanned and recorded by the simulator computers. With the exception of segments on which ATC vectored subjects or instructed altitude changes, exact measurement of performance error was possible by comparison of flight planned route to that actually flown. On normal segments the commanded simulator aircraft position was compared to the flight planned position to provide altitude and course error in feet. Airspeed error was measured in

knots. A composite error index was computed by using the error in each axis relative to the prebriefed limits acceptable to ATC (Table 3-3). Subjects were attempting to remain exactly on course, altitude, and airspeed, but the subjects were additionally warned that ATC would intervene at briefed limits. It was desired that a composite index would be sensitive to deviation relative to both limits, therefore performance error was normalized to one half the prebriefed ATC limits. With an ATC limit of 6000 ft lateral deviation, the composite for course deviation was computed by dividing by 3000 ft. This value was added to airspeed and altitude composite error to create an overall composite error. Composite airspeed error was computed by dividing airspeed error in the cruise phase by 10 knots, or by dividing approach airspeed error by 5 knots. Altitude error was computed by dividing cruise altitude error by 150 feet and approach altitude error by 50 feet.

On segments where ATC commanded changes in a given control axis, the axis changed was not considered as part of performance error. For example, if ATC vectored the subject off of the flight planned route then course error was not included as a part of the composite error measure. Likewise, if the controller directed an altitude change, altitude error was not considered. Airspeed error was always used as part of the composite error measure since the subject always had a reference (command) airspeed.

Average performance information was recorded for each 12 second interval for display. The three 12 second intervals were then averaged to provide a segment average for the given conditions. One average for the normal instrument task and one for the instrument task with the secondary task was recorded for each segment.

4.4. Index of Engagement/Peripheral Temperature Data

Index of Engagement and peripheral temperature data were recorded at a rate of two samples per second, and were also processed using the Basic code (available upon request). Like the performance data, these data were processed continuously, but three consecutive twelve second increments were extracted for the data segments selected during video review. These three increments were also averaged to provide two samples, one normal and one with secondary task, per data segment.

Peripheral temperature data for subjects one through twelve proved to be highly irregular, changing as much as four degrees in a half second and sometimes averaging less than fifty degrees. After review of the video data for peripheral temperature, it was determined that the digital data was unreliable. Video tapes were reviewed again to determine average peripheral temperature for each subject during all data segments. These temperatures were manually substituted into the data records for all subjects.

4.5. Oculometer Data

Oculometer data were gathered as previously described and were processed using the Quick Basic code. The data were recorded in a format providing an X/Y coordinate position on one of six scene planes defined for this study. Only two scene planes, the primary instrument display and the clipboard for distractions, were necessary for analysis of the primary task, however all scene planes were included to minimize loss of fixation data.

Fixations were determined using a combination of saccadic velocity gate and angular size. The velocity gate was one and one-half times the average eye movement

velocity over the previous fifty samples. Fixations were also limited to 1.5 degrees in size which prevented tracking movements from being counted as a single fixation. The start of a fixation was noted any time more than two consecutive data points fell within the angle and velocity criteria stated above. Once a fixation was started, any potential endpoint which exceeded the velocity or angle gates was tested to ensure the point was a valid fixation point. Furthermore, the points immediately following the apparent saccade start point were tested to determine if they fell within the criteria of the previous fixation. If the fixation time was less than the commonly accepted minimum time for a fixation, 100 ms (Viviani, 1995), and at least three points following the apparent end of the fixation fell within the previous fixation criteria, those points were counted as part of that fixation. This prevented extraneous data samples on poor tracking subjects from eliminating fixations.

If the minimum fixation time requirement was met and either the velocity or angle gate criteria was exceeded the subject was considered to be making a saccadic eye movement. If the oculometer lost track an end was declared to the saccade.

The first eye track point of the three necessary to declare the start of a fixation was the point at which saccadic movement was ended. During saccadic movement it is common for the oculometer to lose track due to eye blinks (Stern, et al.,1994). For this study, any loss of track during saccadic eye movement resulted in lost data since no reliable algorithm was found to differentiate between loss of track due to eye blink and loss of track due to other causes.

4.6. Areas of Interest

The location and time for each fixation were recorded with associated factor data. The location for the fixation was the average x/y coordinate location in the scene plane in which the fixation fell. In addition, each scene plane was divided into different areas of interest surrounding the various indicators required to perform the instrument flight tasks. After the x/y location was determined it was compared to a template included in the occulometer code to determine which instrument the subject was viewing.

The boundaries for areas of interest were determined using two methods. First, the boundaries built into the displays were used to divide areas of interest. However, it was noted that all subjects made fixations near the altitude and airspeed indicators without actually entering the area of interest. Part of this phenomenon was due to the accuracy of the oculometer (one degree), but most of these fixations occurred because it was more efficient for the subject to fixate to the inside of the altitude or airspeed indicators where a more rapid return could be made to the attitude indicator. In these cases the areas of interest were expanded toward the attitude indicator using the limits established for normal reading of characters (Rayner and Morris, 1990). Where indicators were too close together to accommodate normal reading limits the areas of interest remained at the physical boundaries.

After each fixation had been place in a specific area of interest the analysis code retained that information to establish the transition pattern to the next fixation. The location of the previous fixation was also recorded with current fixation data. Transition information was important to establishing two other sources of independent variables related to viewing cycles and the transition matrix.

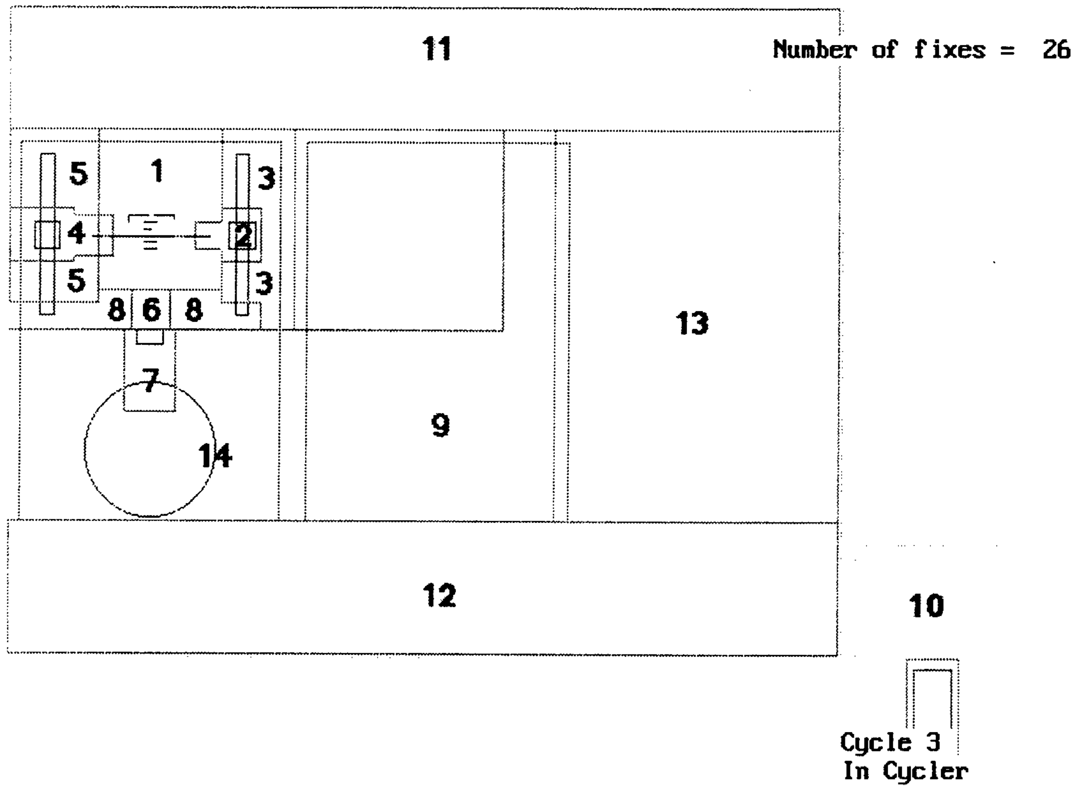

Figure 4.1 Areas of Interest for Transition Analysis

Table 4.1. Display Areas of Interest

Area Number	Display Function
1	Attitude Indicator
2	Airspeed Indicator (Tape)
3	Vertical Velocity Indicator
4	Airspeed Indicator
5	Airspeed Acceleration Indicator
6	Course Indicator Arrow
7	Heading Indicator
8	Course Deviation Indicator
9	Engine Instruments (Upper)/System Displays (Lower)
10	Secondary Task Work Area
11	Out-the-Window Forward Viewing Area
12	Lower Instrumentation Panel (unused for this study)
13	Checklist Display (unused for this study)
14	Horizontal Situation Indicator (HSI)

4.7. Viewing Patterns

It is widely assumed that pilots use some regular, identifiable pattern when performing an instrument flight task. In an attempt to quantify the viewing pattern the altimeter was selected as an anchor from which to measure viewing patterns. This selection was made following initial analysis of results showing altitude was controlled in a similar manner across the subject population. Four measures of the viewing pattern were recorded. The measures were:

1. The number of cycles per 36 second data segment,
2. The average time per viewing cycle,
3. The average number of indicators visited per cycle, and
4. The percentage of useful (performance or control) indicators viewed during a cycle.

Since the cycle count did not start until a fixation fell on the altimeter it was necessary to calculate the average cycle time based on the cycles recorded. The percentage of useful fixations was determined by dividing the number of fixations on instruments necessary for control of flight by the total number of fixations in the cycle. In this manner it was determined what portion of the cycle was used productively, versus scanning portions of the flight deck/windows which provided no pertinent information.

4.8. Transition Matrix

An instrument cross check requires rapid eye movement among areas of interest. Since viewing strategy is an important element of the HIP model, several metrics were based on areas of interest. A 14 x 14 (from - to) transition matrix was built for each 36

second segment of data analyzed. Three independent variables were derived from the transition matrix. The measures were:

1. Percent matrix symmetric,
2. Percent matrix repeat, and
3. Percent matrix useful.

Percentages were used because it was recognized the number of fixations per segment would vary with the quality of the oculometer data. This varied both within and across subjects. Percent matrix symmetric was computed by determining the number of fixations possessing a complimentary transition, and dividing the number by the total number of fixations.

A complimentary transition comes from one of two different sources. First, any time a transition reverses the order (1-4 versus 4-1) it is complimentary. Second, any time consecutive fixations remain within the same area of interest they are considered complimentary. The transition for a fixation remaining within the same area of interest will fall on the diagonal of the transition matrix.

For example, the matrix in Table 4.2 has complimentary fixation transitions between areas 1&4 and 2&4, as well as along the diagonal, but it also has one fixation without a compliment in 1&4 and 2 in 1&3. This matrix would be 9/12, or 75% symmetric. Free viewing, which is unstructured, should result in a nearly symmetric matrix (Ellis, 1986).

Table 4.2. Sample Matrix

Area of Interest	I	II	III	IV
I	2		2	2
II		1	1	1
III			1	
IV	1	1		

The Percent Matrix Repeat considers the number of times the subject stays within the area of interest for more than one fixation. In the above sample matrix four fixations fell on the diagonal (two in Area 1, and one each in areas 2 & 3). The Percent Matrix Repeat would be 4/12 or 33%. Repeat fixations may indicate a specific viewing strategy, or cognitive processing while the subject is fixated on a particular area.

Finally, the Percent Matrix Useful computation was based on the number of fixations which could provide useful information to the subject to perform required activity. Any fixations outside of the primary instruments or secondary task clipboard did not provide information necessary for performing required activity; these fixations were not useful. In addition, fixations dwelling on the same indicator were not always useful since they could have been providing information previously acquired.

Since it was not possible to determine directly if a subject was gaining new information from each fixation a criterion for usefulness was based on the amount of information available within an area of interest. Within a given area of interest there was only one instrument indicator. Some instruments were digital readouts, and some were

analog, however, all were used to determine absolute and trend information. Each instrument provided information when compared to the adjoining legend to determine the absolute value of the indicator. Thus, two fixations could have been required to determine each bit of information. Four fixations could have been required to acquire both absolute and trend information. More than four sequential fixations in any area of interest could not have provided useful information.

The percent "Usefulness" was computed by dividing the number of fixations providing new (useful) information by the total number of fixations for each 12 second segment. The maximum number of sequential fixations in an area of interest providing useful information was four. Any group of sequential fixations exceeding four in one area resulted in an overall loss of "Usefulness". Segments with no fixations were considered to have 0% usefulness.

Chapter 5.

DESIGN OF EXPERIMENT

This study was designed to create a realistic training environment resulting in a variety of performance levels for each subject. Since performance error could result from stress levels related to either task overload or task underload (Yerkes and Dodson, 1908) experimental design created overload (High Load) and underload (Monitor) conditions for each subject based on their performance during simulator familiarization. In addition, a nominal workload condition preceded each high or low workload condition.

A number of factors besides workload affect performance; it is not clear how these same factors affect psychophysiological measures in an aviation environment. Training (Ericsson and Charness, 1994; Klein and Hoffman, 1992) and familiarity with the environment (Blomberg et al, 1993; Lave, 1986) are two factors which have demonstrated a positive correlation with expertise. In addition, if a person is trained or familiar with the environment, then recency is also positively correlated to performance (Lee and Fisk, 1993; Fisk, Lee, and Rogers, 1991).

The addition of a secondary task, circadian rhythm and other factors related to time of day also affect performance (Astrand and Rodahl, 1986). These effects can be either a positive or negative influence on performance depending on the circumstances, but the outcome is usually consistent among subjects for the same conditions.

Although the above factors have documented effects related to performance, their affect on psychophysiological measures in the aviation environment has not been documented for high fidelity simulation of aviation conditions. It is not known whether

the psychophysiological measures employed in this study vary with workload, performance, daytime, secondary task, or some combination thereof. Therefore, the factors above were either controlled or treated in different factor levels described below.

5.1. Workload Levels

Three workload levels were employed in this study. Subjects monitored the autopilot or copilot flying for one quarter of the treatments. This was a low workload condition that was the same for all subjects. The nominal workload condition was a manual flying task under normal conditions. Half of the nominal treatments followed segments in which the subjects were monitoring (low workload), and half of the nominal treatments followed segments in which the subjects had a high workload. Finally, high workload segments constituted one quarter of the segments. Each workload, Monitoring, Nominal (After High Load), Nominal (After Monitoring), and High Load, constituted one quarter of the total segments flown.

5.2. Training

Federal Aviation Regulations provide specific qualification requirements for an individual receiving a pilot instrument rating. An instrument rating is required of all commercial pilots and is a higher level of qualification than a visual flight rules pilot license. Individuals meeting the minimum requirements for an instrument rating may perform worse than people without an instrument rating, but the rating provides a baseline against which training can be measured. This study employed a total of 16 subjects, eight who were instrument qualified and eight who had no pilot rating.

5.3. Familiarity

Familiarity with the handling characteristics, mechanization, and displays for an aircraft type is of such vital importance that pilots must receive special training in the specific aircraft to become "type rated." There is some transfer of training among aircraft types but each aircraft (aircraft simulation) is unique. To determine effect of familiarity on performance and psychophysiological measures, two levels of familiarity were considered. Half of the subjects participating in the study had operated a commercial aircraft or simulator with a glass cockpit. A glass cockpit provides digital instrumentation on a graphical user interface instead of individual analog instruments. The other half of the subjects had no experience on a commercial flight deck, or with a glass cockpit.

Since half of the subjects were rated and half of the subjects were familiar with the cockpit this created a two-by-two matrix of subject pools. One pool, commercial airline pilots, were rated and familiar with the flight deck displays. A second group, Air Force pilots, were instrument rated but not familiar with the glass cockpit. The third group, NASA technicians, were familiar with the glass cockpit but did not possess any aeronautical rating. The fourth group consisted of four subjects who had no aeronautical rating, nor familiarity with a glass cockpit.

5.4. Recency

Recency was controlled within groups. Groups familiar with the commercial flight deck had flown or had simulated flight within the week previous to the study. The groups unfamiliar with the environment had not engaged in any aviation related activity

(except desktop computer flight simulators) for at least the two weeks previous to the simulation study.

Although recency was controlled, different levels of recency provided a confound to training and familiarity. Therefore, recency was grouped with familiarity assuming it would provide the largest effect if either of the factors were significant. Absence of Group significance for performance and psychophysiological measures would indicate these measures were not dependent on familiarity, recency, or training.

5.5. Primary Task and Secondary Task

A secondary task was added to the primary task for half of the data segments. The secondary task was added to provide a second locus of fixations (distraction) away from the primary task to minimize use of automatic scanning strategy. Secondary tasks were designed to require approximately 60 seconds of time for completion. Secondary tasks were mathematical problems requiring addition and multiplication skills.

5.6. Circadian rhythm and Time of Day Factors

All simulations were conducted on the same daily schedule to control variation associated with time of day. Placement of monitoring, nominal, and high load treatments was varied from morning to afternoon to minimize subjects' abilities to learn and predict the simulation scenario. The tendency for subjects to get drowsy after a filling lunch was expected to increase performance error for nominal segments.

Chapter 6.

Performance Results

Results are divided into three chapters. This chapter presents performance results, validates assumptions made in designing the study and analysis code, and creates operationally relevant factor levels for performance error. Chapter 7 presents analysis of eye movement results. Chapter 8 summarizes the peripheral temperature and Index of Engagement results.

Analyses of Variance (ANOVA) conducted consider four factors from the design of experiment. The number of treatments for each factor is below in parenthesis. Each of the 16 subjects provided 64 data segments for a total of 1024 data segments.

Table 6.1. ANOVA Factors and Treatment Levels

FACTOR	TREATMENT LEVELS			
Task Type (2)	Primary Task		Primary and Secondary Task	
Group (4)	Novice	NASA Tech	Military Pilots	Commercial Pilots
Subject (Group)	4	4	4	4
Time of Day (2)	Morning		Afternoon	
Workload Level (4)	Monitor	Nominal (After Monitor)	Nominal (After High Load)	High Load

Root mean squared error is commonly used as the performance standard in aviation studies. This statistic is not operationally relevant for two reasons. First, coupling occurs among control axis so addressing any single axis without consideration of another axis is inaccurate. Second, operational error limits are related to Air Traffic Control (ATC) and safety limits. ATC is gravely concerned if vertical error exceeds

1000 feet, but is totally unconcerned if horizontal error is 1000 feet. The magnitude of acceptable limits varies by control axis and phase of flight.

This chapter provides the results from the methodology developed in Chapter 4 to produce operationally relevant results based on ATC and safety standards. First, the performance error for raw airspeed was analyzed. This same data was then adjusted for performance requirements which varied in a realistic manner between the approach and cruise phases. Second, the resulting ANOVA for adjusted airspeed data is compared to the raw data results to demonstrate the improvement in data quality. The third and fourth sets of ANOVA results are adjusted cross track error and adjusted vertical error. The adjusted cross track error and altitude error results were processed using the same process as the adjusted airspeed results. Next, the three adjusted indices were added together and ANOVA results for a composite index are presented. Adding the three error indices together minimized the effect of subjects who emphasized one axis of control over another. Finally, data points from the composite error index were translated into the four error ratings of practical significance to the subjects.

The error ratings of practical significance to subjects were: 1) No error, 2) Error within tolerances prescribed by Air Traffic Control (ATC), 3) Error exceeding ATC tolerances but within safe limits, and 4) Error great enough to pose safety concerns due to mid-air collision or impact with the ground. After the composite error index was translated into error ratings an ANOVA was conducted on the Performance Error Ratings.

6.1. Study Validation

The first goal of this study was to create an aviation performance model relating performance to workload levels. To accomplish this goal, the designed Workload Levels had to produce variation in performance. It was also desirable to produce error serious enough to pose safety concerns. It was hoped this anomalous performance could be related to psychophysiological measures.

6.2. Dependent Variables

The study was designed to produce various levels of airspeed, altitude and course performance error by variation of workload. This performance error was produced by requiring subjects to perform instrument flight tasks while varying workloads directly related to the instrument flight task. Increases in the task level of difficulty were expected to result in an increase in performance error. All six error measures presented below varied significantly with workload, $F(12,3) = 9.24\text{-}141.65$, $p < 0.001$.

Although the second and third Workload Levels were the same workload, they were considered as separate treatments. They were considered as such, since it was expected performance would differ on tasks of moderate difficulty following difficult workload level versus following easy workload. Variance among the performance variables was not the same in all four treatments. The easiest workload level (Monitor treatment) was the baseline condition in which the subjects were monitoring instrument parameters. By definition, performance error was zero in this treatment since the subjects were not controlling the simulator. Although ANOVA included only three Workload Levels the baseline condition was added in figures for the sake of comparison. The

Nominal treatments were expected to yield low levels of performance error. The High Load workload level (treatment four) was expected to yield performance ranging from nominal to unsatisfactory. Thus, variance of the raw error was expected to increase with the Workload Level. Airspeed error was analyzed first since it was measured for all segments. Both altitude and cross track error were not measured for two segments on which ATC was directing changes in the given axis.

6.2.1. Airspeed Error

Table 6.2 displays results from the five factor ANOVA for airspeed error. Airspeed was measured in nautical miles per hour (knots). Subjects exceeded the prescribed ATC limits for airspeed error in 152 of 1024 data segments.

Knots of Indicated Airspeed (KIAS) was not manipulated to affect the difficulty of the instrument flight task. KIAS limits did vary appropriately between approach phase and cruise phase. Figure 6.1 demonstrates error increased with Workload Level, $F(2,24) = 33.75$, $p < 0.001$. This figure uses bold lines and arrows to indicate a counterclockwise progression of the data points. This progression indicates the Nominal (After High Load) treatment had a higher average value than the Nominal (After Monitor) treatment. When the Nominal (After Monitor) value is higher, clockwise progression is indicated by arrows and a normal, thinner (not bold) line. The difference is apparent in Figure 6.2.

Table 6.2. Five Factor ANOVA for Airspeed Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	219.34	219.34	T x S(G)	2.23	NS
Group	[G]	3	249.16	83.05	S(G)	0.49	NS
#Subject(Group)	[S(G)]	12	8098.22	674.85		4.02***	
Time Day	[D]	1	0.13	0.13	D x S(G)	0.00	NS
Workload Level	[W]	2	4484.86	2242.43	W x S(G)	33.75***	0.202
Interactions							
T x G		3	292.49	97.50	T x S(G)	0.99	NS
T x D		1	5.92	5.92	T x D x S(G)	0.29	NS
T x W		2	79.67	39.83	T x W x S(G)	0.97	NS
G x D		3	135.72	45.24	D x S(G)	0.43	NS
G x W		6	246.52	41.09	W x S(G)	0.62	NS
D x W		2	3013.12	1506.56	D x W x S(G)	18.09***	0.132
T x G x D		3	155.67	51.89	T x D x S(G)	2.52	NS
T x G x W		6	380.86	63.48	T x W x S(G)	1.55	NS
T x D x W		2	768.46	384.23	T x D x W x S(G)	8.45**	0.032
G x D x W		6	707.23	117.87	D x W x S(G)	1.42	NS
T x G x D x W		6	334.75	55.79	T x D x W x S(G)	1.23	NS
Error Terms							
S(G)		12	2024.56	168.71			
T x S(G)		12	1179.98	98.33			
D x S(G)		12	1258.12	104.84			
W x S(G)		24	1594.45	66.44			
T x D x S(G)		12	247.27	20.61			
T x W x S(G)		24	985.79	41.07			
D x W x S(G)		24	1998.58	83.27			
T x D x W x S(G)		24	1091.40	45.48			
Total		191	21454.05				

* = p<0.05   ** = p<0.01   *** = p<0.001   NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

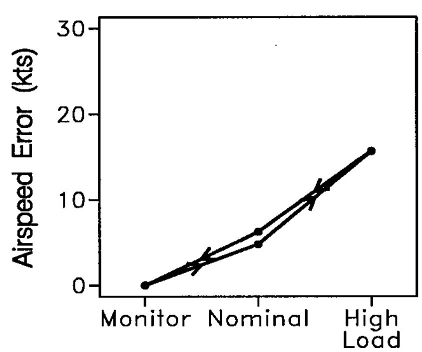

Figure 6.1. Airspeed Error versus Workload Level

The nominal treatments were not significantly different from each other. However their averages were significantly different from the Monitor, $p < 0.01$, and High Load, $p < 0.001$, treatment levels. The remaining ANOVA factors, Task, Group and Time of Day, were not significant in accounting for airspeed error variance. Statistical significance of the interactions below was determined using Tukey's Test.

Table 6.3. Factor Level Data for Airspeed Error

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	NA	6.23	4.81	15.70
Std Dev	NA	12.47	9.74	19.88
Time of Day	*AM*		*PM*	
Average	8.94		8.89	
Std Dev	16.25		14.60	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	8.15	7.83	10.77	8.90
Std Dev	12.88	16.30	16.90	15.30
Task	*Primary*		*Primary with Secondary*	
Average	9.98		7.85	
Std Dev	16.39		14.34	

Two interactions reached significance. The first significant interaction was between Time of Day and Workload Level, $F(2,24) = 18.09$, $p < 0.001$. This interaction was expected from the design of experiment. A study by Wickens et al (1995) indicated the airspeed error was most sensitive to the instrument flight learning process. Subjects attended first to problems with the spatially coupled variables, altitude and course. As additional attention resources become available from efficiency gained in training more attention was devoted to airspeed. Thus, airspeed error would decrease most as a result of any training effects exhibited.

Figure 6.2, illustrates how the high load treatment resulted in significantly more performance error in the morning, $p < 0.001$, compared to all the afternoon treatments. All subjects performed better on the PM/High Load treatment compared to the morning.

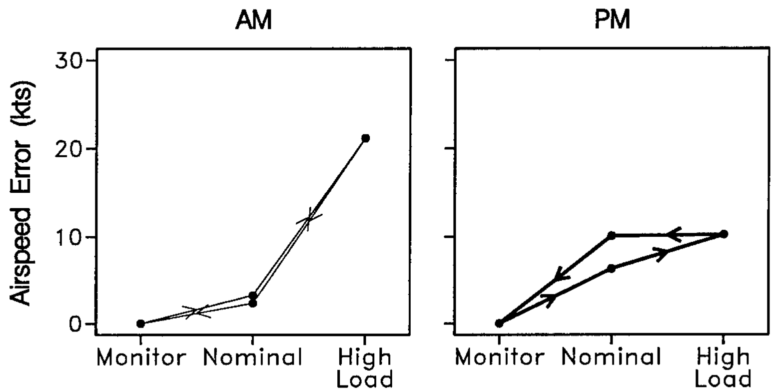

Figure 6.2. Airspeed Error interaction of Time of Day /Workload Level

The other remaining interaction was a three factor interaction involving Task, Time of Day, and Workload Level, $F(2,24) = 8.45$, $p < 0.01$. Neither morning nor afternoon performance exceeded ATC restrictions on the average. The more experienced aviators showed a greater performance decrement for Nominal After High Load segments, but not as great for the Nominal After Monitor segments. In the afternoon, a reduction in workload was accompanied by a reduction in performance while in the morning this was not true. The afternoon decrement was expected since the study's design attempted to exploit a tendency for drowsiness after lunch. The upper right panel of Figure 6.3, Afternoon/Primary Task (PT), illustrates the difference in this treatment. This same treatment will be the source of significant interactions for numerous

psychophysiological factors. The lower panels possess data for Primary Task and Secondary Task (PT + ST) treatments.

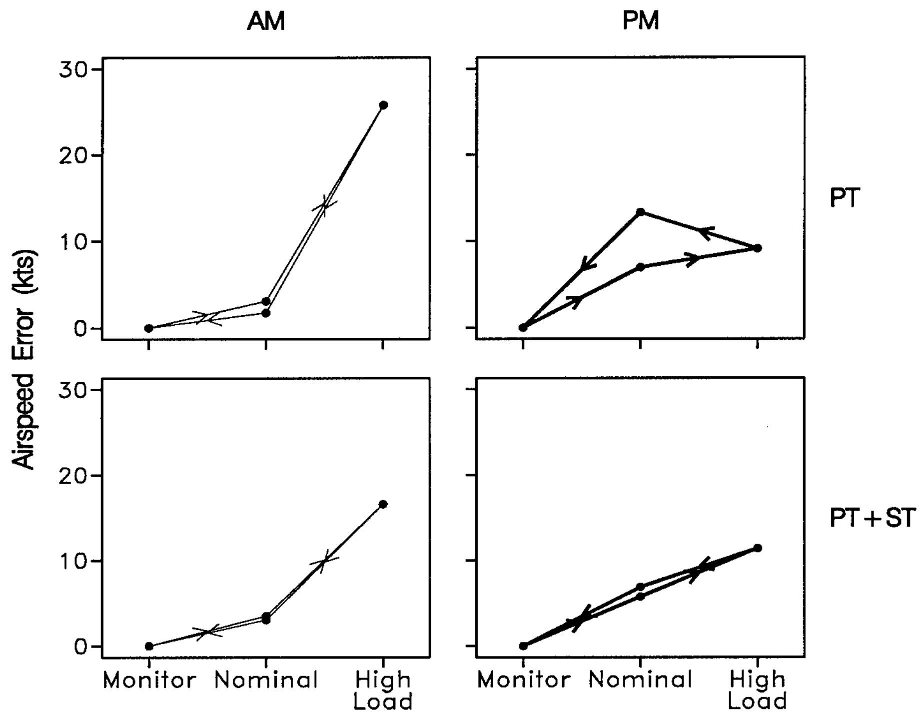

Figure 6.3. Airspeed Error interaction of Task /Time of Day/Workload Lvl

The AM/High Load treatment exhibited significantly greater error than all the other results, $p < 0.001$, but in an unexpected manner. The Morning/High Load treatments exhibited greater error than all other treatments in the morning. There was no significant difference among Monitor and Nominal treatments in the morning. Normally omission of the secondary task would be expected to reduce error by allowing greater attention on the primary task.

In summary, one treatment and two interactions were statistically significant for airspeed error. Very few performance deviations in airspeed error were practically

significant relative to the maximum ATC error limit of 20 Knots. However, some errors occurred on approach legs, where the ATC limit was 10 Knots. These results would have been highlighted by ATC if adjusted for phase of flight. In the raw airspeed ANOVA, 36.8% of variance was accounted for (ω^2) by the significant factors and interactions.

6.3. Adjusted Performance Indices

Altitude, course, and airspeed error provide basic input for performance metrics. Aviation performance is not an absolute scale. Airspeed of an aircraft at 40,000 feet on an intercontinental route is limited most by thrust efficiency and aircraft structural integrity. However, in the instrument pattern used in this study, tighter limits are required to maintain aircraft separation. On final approach, even tighter airspeed limits are necessary for safety. By "adjusting" airspeed error relative to ATC limits, the limits briefed to subjects become the yardstick against which their performance is measured. These limits mirror those used operationally, although the exact numbers vary somewhat by aircraft model and air traffic control area. This same approach to "adjusting" error was used to consider Cross Track Error and Altitude Error.

6.3.1. Adjusted Airspeed Error

Subjects were briefed that airspeed deviations of more than 20 Knots on cruise legs would bring admonition from the controller. In addition, the co-pilot would warn subjects if they were more than 10 Knots off of programmed airspeed while on approach. More than 10 Knots error on approach airspeed could result in stalling the aircraft if slow or running off of the runway if fast.

Data was adjusted for the different ATC limits placed on airspeed error for different segments. The adjusted index equalized the severity of error relative to these limits. To create sensitivity to the magnitude of error, the absolute airspeed error was divided by one-half of the ATC limit briefed to subjects. Thus, a 10 Knot error in cruise phase (20 Knot error limit) would result in an adjusted error index of one. The same error in the approach phase (10 Knot error limit) would result in an adjusted error index of two. After adjustment, an index rating of two would place the simulator at the ATC limit. If airspeed error increased beyond this limit, ATC intervened.

All effects found in the ANOVA for Airspeed Error were also present in the five factor ANOVA for Adjusted Airspeed Error. Other factors approached, but did not reach significance with the adjusted index. The total amount of variance accounted for by significant factors also increased. Table 6.3 contains the ANOVA results. As with the Airspeed Error ANOVA, Workload Level, $F(1,12) = 25.76$, $p < 0.001$, was significant. While reviewing these results, it should be noted that the ATC level for error significance is two. This level would represent either 10 Knots error in approach or 20 Knots error in cruise. The relationship between Adjusted Airspeed Index and Workload Level is the same as previously presented with Airspeed Error. The results of Adjusted Airspeed Index are slightly more significant (total $\omega^2 = 43.1\%$) due to a reduction in variation. Despite reduction in variation, data displayed heteroscedasticity for all significant factors.

Subjects exceeded the prescribed ATC limits for adjusted airspeed error in 181 of 1024 data segments. A breakdown of the treatments in which errors occurred and the severity of the error will be presented with composite performance variables.

Table 6.4. Five Factor ANOVA for Adjusted Air Speed Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	11.72	11.72	T x S(G)	3.29	NS
Group	[G]	3	13.92	4.64	S(G)	0.94	NS
#Subject(Group)	[S(G)]	12	236.67	19.72		3.83***	
Time Day	[D]	1	1.27	1.27	D x S(G)	0.41	NS
Workload Level	[W]	2	121.80	60.90	W x S(G)	25.76***	0.140
Interactions							
T x G		3	11.11	3.70	T x S(G)	1.04	NS
T x D		1	0.18	0.18	T x D x S(G)	0.23	NS
T x W		2	4.16	2.08	T x W x S(G)	1.68	NS
G x D		3	7.07	2.36	D x S(G)	0.75	NS
G x W		6	6.88	1.15	W x S(G)	0.49	NS
D x W		2	228.40	114.20	D x W x S(G)	38.62***	0.265
T x G x D		3	4.16	1.39	T x D x S(G)	1.76	NS
T x G x W		6	11.54	1.92	T x W x S(G)	1.55	NS
T x D x W		2	25.20	12.60	T x D x W x S(G)	6.51**	0.025
G x D x W		6	20.69	3.45	D x W x S(G)	1.17	NS
T x G x D x W		6	14.83	2.47	T x D x W x S(G)	1.28	NS
Error Terms							
S(G)		12	59.17	4.93			
T x S(G)		12	42.77	3.56			
D x S(G)		12	37.68	3.14			
W x S(G)		24	56.73	2.36			
T x D x S(G)		12	9.47	0.79			
T x W x S(G)		24	29.72	1.24			
D x W x S(G)		24	70.97	2.96			
T x D x W x S(G)		24	46.48	1.94			
Total		191	835.91				

* = p<0.05   ** = p<0.01   *** = p<0.001   NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

In summary, the level of significance increased marginally when airspeed error was adjusted to appropriate ATC limits for the segment flown. Variance within factors was reduced, but variance was still not homogenous, $p < 0.05$. The amount of variance for which the ANOVA accounted increased from 36.8% for Airspeed to 43.1% with Adjusted Airspeed. Type and form of significant factors and interactions did not change.

6.3.2. Adjusted Cross Track Error

Table 6.5 contains results from the five factor ANOVA for Adjusted Cross Track Error. Cross track error was measured in feet from the planned course line. Adjusted Cross Track Error was created by dividing absolute cross track error by one-half the briefed ATC limit for the segment. Like Adjusted Airspeed Error, the difference

Table 6.5. Five Factor ANOVA for Adjusted Cross Track Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	148.55	148.55	T x S(G)	71.16***	0.089
Group	[G]	3	4.57	1.52	S(G)	0.35	NS
#Subject(Group)	[S(G)]	12	206.22	17.19		1.26	
Time Day	[D]	1	74.49	74.49	D x S(G)	19.76***	0.043
Workload Level	[W]	2	185.88	92.94	W x S(G)	25.46***	0.108
Interactions							
T x G		3	11.62	3.87	T x S(G)	1.86	NS
T x D		1	41.62	41.62	T x D x S(G)	15.60**	0.024
T x W		2	116.89	58.45	T x W x S(G)	23.00***	0.068
G x D		3	2.51	0.84	D x S(G)	0.22	NS
G x W		6	2.90	0.48	W x S(G)	0.13	NS
D x W		2	293.50	146.75	D x W x S(G)	40.78***	0.174
T x G x D		3	9.24	3.08	T x D x S(G)	1.15	NS
T x G x W		6	21.41	3.57	T x W x S(G)	1.40	NS
T x D x W		2	237.56	118.78	T x D x W x S(G)	45.23***	0.141
G x D x W		6	19.08	3.18	D x W x S(G)	0.88	NS
T x G x D x W		6	21.11	3.52	T x D x W x S(G)	1.34	NS
Error Terms							
S(G)		12	51.56	4.30			
T x S(G)		12	25.05	2.09			
D x S(G)		12	45.23	3.77			
W x S(G)		24	87.60	3.65			
T x D x S(G)		12	32.01	2.67			
T x W x S(G)		24	60.99	2.54			
D x W x S(G)		24	86.37	3.60			
T x D x W x S(G)		24	63.03	2.63			
Total		191	1642.78				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

in cross track error was significant for Workload Level, $F(2,24) = 25.46$, $p<0.001$. In addition, Time of $F(1,12) = 19.76$, $p < 0.01$ and Task, $F(1,12) = 71.16$, $p < 0.001$ were significant. Group was not significant. Subjects exceeded the prescribed ATC limits for cross track error in 104 of 1024 data segments.

Addition of the secondary task coincided with a significant decrease in Cross Track Error, $p < 0.001$. The average level of error with only the primary task was significant by ATC standards, but was within ATC standards with the secondary task. Subjects had also performed marginally better with a secondary task for Adjusted Airspeed Error.

Much like airspeed error, cross track error increased slightly from the Monitor treatment to the Nominal treatments and then increased significantly for the High Load treatments, $p < 0.001$. Again, there was no significant difference between the two nominal treatments. There was also no statistical difference among the two nominal treatments and the Monitor treatment. However, the High Load treatment was statistically different from the other three treatments, $p < 0.001$. The average Adjusted Cross Track Error for High Load treatments exceeded the ATC briefed limits.

Four interactions resulted from the five factor analysis of variance on composite cross track error. Two two-way interactions involved Task. The first was a two factor interaction between Task and Time of Day. The morning simulation segments with only a primary task were statistically higher than the three other results, $p < 0.01$. The Primary Task/Morning treatment also exceeded ATC standards. The overall error level decreased for the afternoon results. Elevated error levels with only Primary Task were not expected but were a consistent result of the experimental design.

In the Task/Workload Level interaction, the Primary Task\High Load treatment was different from all other results. There was a regular increase in cross track error with increase in Workload Level in the presence of both a primary and secondary task. However, the increases were so small that no significant difference existed among these treatments when the secondary task was added. Increases in error are evident, $p < 0.001$ in Nominal After High Load and High Load treatments without a secondary task. Variance increases in both these conditions as well.

Subjects had the option of performing the secondary task at their discretion as long as it was completed on the specified segment. In general, subjects chose to complete the task when they perceived they had minimized error for their primary task. Thus, performance error measured with the secondary task represents a minimum error level for the primary aviation task. This is a consistent result among performance indices.

Another two factor interaction occurred between Time of Day and Workload Level. The High Load treatment in the morning resulted in a level of error statistically greater than all other data points, $p < 0.001$. The other treatment causing the interaction was the afternoon Nominal After High Load treatment. This treatment was statistically different from all other treatments, $p < 0.05$, except the morning Nominal After Monitor treatment.

As previously stated, the study was designed to produce a loss of vigilance resulting in an elevated level of error during the afternoon. This loss of vigilance manifested itself in the Nominal After High Load/PM treatment where subjects perceived a workload reduction following the High Load conditions. Among the afternoon

treatments, performance on the Nominal After High Load Workload Level was most similar to the High Load treatment. This same treatment also showed elevated performance error for the same interaction in airspeed error.

Finally, a three factor interaction among Task, Time of Day, and Workload Level occurred like the same interaction previously shown with Airspeed Error. (The autopilot and co-pilot also maintained the simulation in a low performance state.) By definition, there was no error for the Monitor treatments since subjects were not controlling the simulator at the time. However, results from the Nominal After High Load and High Load treatments with only a primary task were high, $p < 0.001$, compared to all other treatments. At the Nominal After High Load treatment, error was low in the morning but increased significantly, $p < 0.001$, in the afternoon. The opposite occurred at the High Load treatment. Thus far, every result presented has revealed elevated performance error for the High Load morning segments with only primary task and for the Nominal After High Load afternoon segments with only primary task.

In summary, the significant factors and interactions seen in Adjusted Airspeed Error were also present and similar in direction to Adjusted Cross Track Error. In addition, a Task/Time of Day interaction was present for Adjusted Cross Track Error.

6.3.3. Adjusted Vertical Performance Error

Table 6.6 contains results from the five factor ANOVA for Adjusted Vertical Error. Vertical error was measured in feet from the planned course line. Adjusted Vertical Error was created by dividing absolute vertical error by one-half the briefed ATC limit for the segment. Like Adjusted Airspeed Error and Adjusted Cross Track

Error, the difference in error was significant for Workload Level, $F(2,24) = 9.24$, $p < 0.01$, and Task, $F(1,12) = 11.65$, $p < 0.01$. Consistent with the Adjusted Airspeed Error the factors of Group, Subject, and Time of Day were not significant. Subjects exceeded the prescribed ATC limits for vertical error in 225 of 1024 data segments. Variance was not homogeneous for any factor.

Table 6.6. Five Factor ANOVA for Adjusted Vertical Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	18.85	18.85	T x S(G)	11.65**	0.013
Group	[G]	3	11.30	3.77	S(G)	1.09	NS
#Subject(Group)	[S(G)]	12	166.16	13.85		1.02	NS
Time Day	[D]	1	8.97	8.97	D x S(G)	2.70	NS
Workload Level	[W]	2	34.81	17.40	W x S(G)	9.24**	0.023
Interactions							
T x G		3	20.23	6.74	T x S(G)	4.17*	0.012
T x D		1	69.15	69.15	T x D x S(G)	34.34***	0.051
T x W		2	122.49	61.24	T x W x S(G)	26.60***	0.089
G x D		3	1.53	0.51	D x S(G)	0.15	NS
G x W		6	7.45	1.24	W x S(G)	0.66	NS
D x W		2	535.76	267.88	D x W x S(G)	112.72***	0.401
T x G x D		3	3.84	1.28	T x D x S(G)	0.64	NS
T x G x W		6	14.64	2.44	T x W x S(G)	1.06	NS
T x D x W		2	97.56	48.78	T x D x W x S(G)	24.32***	0.071
G x D x W		6	23.11	3.85	D x W x S(G)	1.62	NS
T x G x D x W		6	20.67	3.44	T x D x W x S(G)	1.72	NS
Error Terms							
S(G)		12	41.54	3.46			
T x S(G)		12	19.42	1.62			
D x S(G)		12	39.83	3.32			
W x S(G)		24	45.20	1.88			
T x D x S(G)		12	24.16	2.01			
T x W x S(G)		24	55.25	2.30			
D x W x S(G)		24	57.04	2.38			
T x D x W x S(G)		24	48.13	2.01			
Total		191	1320.94				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Task was significant, $p<0.01$, but not to the level found for the other two adjusted indices ($p < 0.001$). Like the other adjusted error indices, the level of error decreased when the secondary task was added. The average vertical error did not exceed the ATC briefed limits with or without addition of a secondary task. Cross track error exceeded ATC limits with primary task. This factor was not significant for airspeed error performance.

Workload Level was a significant factor. As with the other two adjusted indices the Nominal treatments were not significantly different from each other. However, the Nominal After High Load was again slightly higher than the Nominal After Monitor treatment. Together they were significantly different from both the Monitor and High Load treatments, $p<0.001$. The High Load treatment resulted in an average (2.75) exceeding the ATC index limit of two.

Four interactions were significant for adjusted vertical error. First, the factors of Task and Time of Day again showed a significant interaction $F(1,12) = 34.34$, $p < 0.001$. In this case, the interaction was different from the interaction in Adjusted Cross Track Error. In Adjusted Vertical Error the primary with secondary task error increased for the afternoon, while it was flat for the Adjusted Cross Track Error.

The second interaction occurred between Task and Workload, $F(2,24) = 26.6$, $p < 0.001$. The interaction has exactly the same trends in this interaction for the other two adjusted indices. There was a relatively smooth progression of error with the addition of a secondary task. The High Load treatment without the secondary task was significantly greater than all other results, $p < 0.001$. The Nominal After High Load treatment differed

from the other results with no secondary task, $p < 0.05$. As with the other adjusted indices, variance was not homogenous for the treatments without a secondary task.

The third, two factor interaction was between Time of Day and Workload. Again, there was a distinctive peak for the Nominal After High Load treatment in the afternoon. This peak was significantly greater, $p < 0.001$, than the Monitor and Nominal treatments in the morning. The AM/High Load was significantly different, $p < 0.001$, from all results except the Nominal After High Load treatment in the afternoon.

Finally, there was the same three factor interaction in the adjusted vertical error ANOVA as was found in the other two adjusted indices, the interaction among Task, Time of Day, and Workload Level factors. The relationship between Task and Time of Day was different for each Workload Level. All results were the same for the Monitor treatment. For the Nominal After Monitor treatment, the afternoon condition with a secondary task differed from the other three levels of difficulty, $p < 0.01$. This treatment is the only instance where error increased with the addition of a secondary task. Unlike previous control axis, there was no Primary Task/PM/Nominal After High Load interaction.

In summary, like the results from the other adjusted indices, Adjusted Vertical Error varied significantly with Task and Workload Level. The presence of a secondary task reduced the variation in results and in one case, increased performance error. This error increase was not seen in the other adjusted indices. There were clearly some changes in control strategies from morning to afternoon, as evidenced by results for the adjusted error indices becoming more similar in the afternoon.

6.3.4. Composite Performance Error

Since there was evidence of changing control strategies and the possibility that different subjects might put different priorities on the three control axis, a Composite Performance Error index was created. This composite index adds the adjusted error from all three performance axis together to minimize variation due to changing priorities.

Different control axis required different control strategies since control input methodology and aircraft response time are unique for each axis. Use of any single axis performance error index would be confounded by the subjects' shifts in control strategy as they attempt to find an optimal control strategy.

Table 6.7 contains factor means and standard deviations, while results from a five factor ANOVA for Composite Performance Error are in Table 6.7. Composite

Table 6.7. Factor Level Data for Composite Performance Error

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	N/A	3.80	2.70	9.15
Std Dev	N/A	5.54	2.95	11.70
Time of Day	***AM***		***PM***	
Average	5.03		5.40	
Std Dev	10.57		4.65	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	5.06	4.60	5.58	5.64
Std Dev	6.86	7.37	8.75	9.43
Task	***Primary***		***Primary with Secondary***	
Average	6.60		3.84	
Std Dev	10.73		3.79	

Performance Error was computed by adding together the three adjusted error indices for each the 1024 different conditions. Both Task, $F(1,12) = 30.78$, $p < 0.001$, and Level of

Workload, $F(2,12) = 50.27$, $p < 0.001$, produced significant differences among treatments. There was no significant difference among treatments for Group, or Time of Day. Since each of three control axis could contribute up to two units of performance error without exceeding ATC standards, the ATC tolerance level on the index was six. Performance error exceeded ATC tolerances on 211 of 1024 segments analyzed. Some 63 segments had performance error exceeding ATC tolerances on two of three adjusted indices, but no segments exceeded tolerances on all three indices.

Table 6.8. Five Factor ANOVA for Composite Performance Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	34.92	34.92	T x S(G)	6.85*	0.023
Group	[G]	3	6.64	2.21	S(G)	0.27	NS
#Subject(Group)	[S(G)]	12	394.06	32.84		4.31***	
Time Day	[D]	1	3.36	3.36	D x S(G)	0.62	NS
Workload Level	[W]	2	472.12	236.06	W x S(G)	71.63***	0.358
Interactions							
T x G		3	16.68	5.56	T x S(G)	1.09	NS
T x D		1	1.25	1.25	T x D x S(G)	0.80	NS
T x W		2	17.11	8.56	T x W x S(G)	4.44*	0.010
G x D		3	2.48	0.83	D x S(G)	0.15	NS
G x W		6	13.59	2.27	W x S(G)	0.69	NS
D x W		2	39.86	19.93	D x W x S(G)	4.47*	0.024
T x G x D		3	5.27	1.76	T x D x S(G)	1.12	NS
T x G x W		6	23.44	3.91	T x W x S(G)	2.03	NS
T x D x W		2	50.29	25.15	T x D x W x S(G)	9.56***	0.035
G x D x W		6	52.52	8.75	D x W x S(G)	1.97	NS
T x G x D x W		6	17.48	2.91	T x D x W x S(G)	1.11	NS
Error Terms							
S(G)		12	98.52	8.21			
T x S(G)		12	61.16	5.10			
D x S(G)		12	64.54	5.38			
W x S(G)		24	79.10	3.30			
T x D x S(G)		12	18.79	1.57			
T x W x S(G)		24	46.28	1.93			
D x W x S(G)		24	106.91	4.45			
T x D x W x S(G)		24	63.16	2.63			
Total		191	1295.48				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Of the 211 segments exceeding the ATC tolerances, 18 of these errant segments could be considered life threatening. In these 18 segments, performance was so poor that a mid-air collision with a nearby aircraft or ground collision was possible. ATC tolerances are designed such that two aircraft could be at the ATC limit, encroaching on each other and a significant buffer would still exist. Aircraft would have to err at four times the ATC limit (or eight index units) on one of the three axis to be life threatening.

The composite effect of secondary task on performance error was consistent with all adjusted error indices previously presented. Performance of the instrument flight task with a secondary task was consistently better than that without the secondary task.

It should be noted that subjects were presented with the secondary task immediately after rolling out of a turn onto a new leg of the instrument pattern. Subjects were not instructed to complete the task immediately, they were instructed to complete the task as soon as possible, but before the end of the leg onto which they had just turned. With knowledge of the task requirements for the segment, subjects used their judgement to decide when it was best to complete the secondary task. Subjects attempted to minimize their performance error before attempting the secondary task.

Subject was a significant factor for only the airspeed error adjusted index, but it was also a significant factor for the composite index. In the case of the Adjusted Airspeed Index, the average levels of error did not reach the ATC briefed limits. Only the best performing subject (6) and worst performing subject (16) were different, $p < 0.05$.

The Time of Day factor was marginally significant for the Cross Track Error Index. Although Time of Day was not significant for the Airspeed or Vertical Error

indices, the small performance changed which did occur were in the opposite directions for the indices from morning to afternoon. Thus, when the indices were added together the Composite Error Index was not significantly different for Time of Day.

There were significant differences for performance among the Workload Level treatments displayed Figure 6.4. As with the individual indices there was no significant difference between the Nominal Workload Levels. These two levels were the same task. The nominal segments were both significantly different, $p < 0.001$, from both the Monitor and High Load treatments.

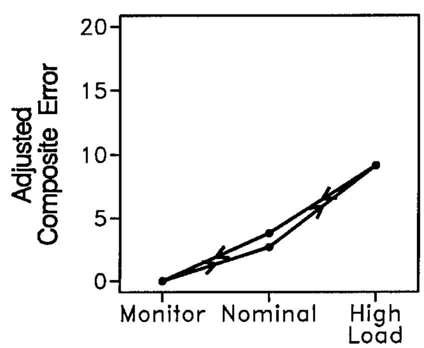

Figure 6.4. Adjusted Composite Performance Error versus Workload

There were four, two factor interactions resulting from the ANOVA for Composite Performance Error. Task interacted with Time of Day. Morning segments with a primary and secondary task resulted in significantly better performance than those with only the primary task, $p < 0.001$. The Primary with Secondary/AM treatment was also better than the two afternoon treatments, $p < 0.01$. Across the indices the data from the primary with secondary task showed consistently better performance.

The second interaction, between Task and Workload Level was consistent across the adjusted indices. There was no Workload Level at which addition of a secondary task resulted in decreased performance. In fact, the opposite was true for both Nominal After High Load, $p < 0.05$, and High Load, $p < 0.001$, treatments. Addition of the secondary task produced a smooth progression for the minimum level of performance error.

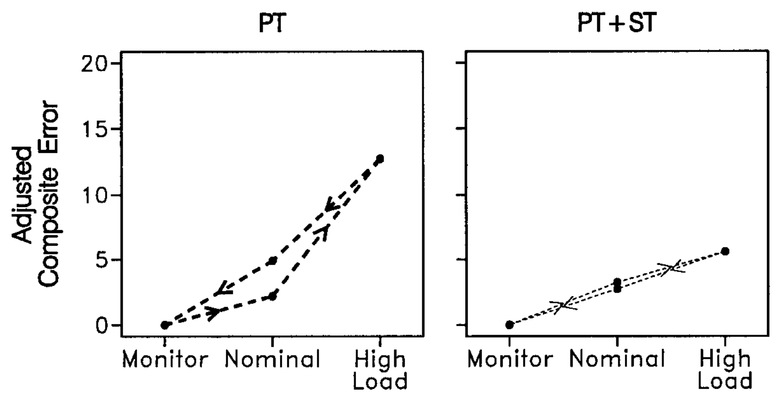

Figure 6.5. Composite Performance Error interaction of Task/Workload Level

The third interaction, between Time of Day and Workload Level, was also present in all adjusted indices. As expected, the High Load treatment in the morning resulted in the worst average performance results, $p < 0.001$. Another expected result was the interaction in the afternoon for the Nominal After High Load treatment. This is illustrated by the peak in performance error for the Nominal After High Load treatment (Figure 6.6). The workload was the same as with the Nominal After Monitor treatment. This coincided with a reduction in workload because it followed the High Load treatments. The afternoon segments, which followed a generous lunch with no caffeine, were those on which a loss of vigilance was expected. The afternoon results for the

Nominal After High Load treatment were significantly different from all morning results, $p < 0.001$. In addition, they were significantly different from other afternoon results for the Monitor and Nominal After Monitor treatments, $p < 0.001$. They were not significantly different from the High Load treatment.

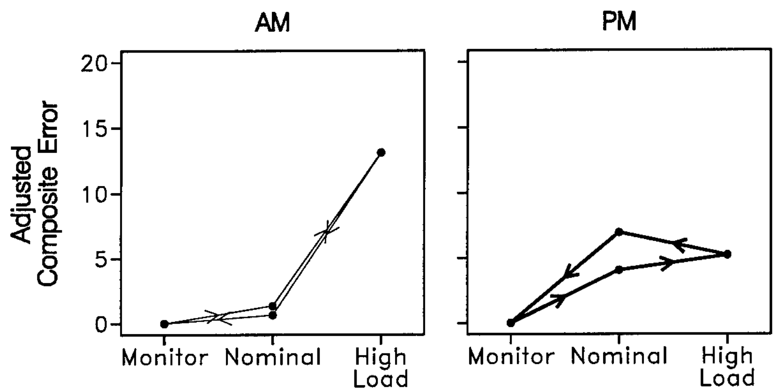

Figure 6.6. Composite Performance Error interaction of Time of Day/Workload Level

All three adjusted indices displayed similar trends for the afternoon results at the Nominal After High Load treatment. The magnitude and statistical significance of the changes varied somewhat. However, results from the High Load treatment in the afternoon were never statistically different from the Nominal After High Load treatment in the afternoon.

Finally, there was a three factor interaction among the factors of Task, Time of Day, and Level of Workload. Results are displayed in Figure 6.7. If an increase in performance error was due to a decrease in afternoon vigilance, then a three factor interaction would indicate a shift in strategy between morning and afternoon for the same workload. The viewing strategy change would most likely manifest itself for segments in

which workload decreased. Those segments would be the Monitor treatments following Nominal Workload Levels and the Nominal After High Load treatments.

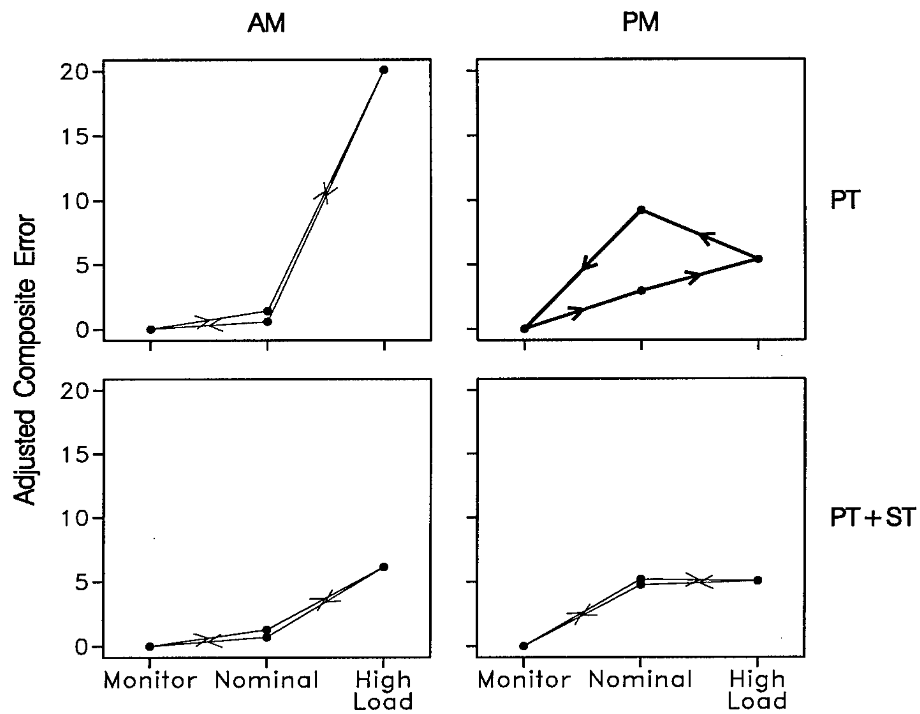

Figure 6.7. Composite Performance Error interaction of Task/Time of Day/Workload

No recorded performance drop was possible for the Monitor treatments since subjects were only monitoring. However, the indicated result was found for the Nominal After High Load treatments among the three individual indices and in the composite index.

The Nominal After Monitor treatment showed no significant interaction with either Time of Day or Task factors. However, the Nominal After High Load treatment interacted to show a significant increase in performance error, $p < 0.01$, for afternoon treatments with and without a secondary task.

6.3.5. Summary of Composite Performance

In summary, any marginal significance found in Time of Day and Group disappeared when adjusted indices were combined. This leveling of error across time of day and subject groups supports the contention that subjects reprioritized control axis to optimize their control strategies. Total error between the morning and afternoon was consistent. Task and Workload Level factors remained significant.

The percent of variance accounted for by the specified ANOVA factors increased 4.3% by going to an adjusted index model. The airspeed index accounted for 0.431 of variance. The cross track error index accounted for 0.649 of variance, and the vertical error index accounted for 0.560 of variance. The total variance accounted for in the composite index was 0.632. The two greatest sources of variation for all indices were Workload Level and the interaction between Time of Day/Workload Level. Variance was not homogeneous for any of the ANOVA previously discussed. A Greenhouse-Geisser (1959) approach to analysis, which accounts for variance effects on significance of results, revealed that all significant factors remained significant. In fact, only the three factor interaction from the Adjusted Airspeed Error Index decreased in significance level from $p < 0.01$ to $p < 0.05$.

6.4. Performance Error Rating.

An error rating was developed in an effort to remove heteroscedasticity and to provide a basis for analysis of psychophysiological variables as they relate to levels of performance error. Many data segments had no measurable error and others having error were well within ATC standards. Other performance errors exceeding ATC criteria

caused some concern but resulted in no real danger since ATC limits are conservative. Finally, workload and stress level would undoubtedly increase, if the performance was so poor that the subject was in danger of crashing into the ground or another aircraft. One of the hypotheses of this study was that different operationally significant levels of error would result in different psychophysiological measures. Performance Error Rating translates results of the Composite Error Index into a scale from zero to three where zero was low error and three was dangerous error.

Composite Error Index results with low error (less than two on the Composite Error Index) were rated zero. Measurable error within ATC tolerances (less than six on the Composite Error Index) were rated one. Performance errors exceeding ATC criteria but not in danger of crashing were rated two, and dangerous performance error (Composite Error Index greater than 24) was rated three. These ratings are based on operationally recognized performance limits.

6.4.1. Performance Error Rating

The five factor ANOVA for Performance Error Rating accounted for 73% of the variance. Variance was homogeneous across all significant factors. Without the Monitor treatments (observation segments) data was normally distributed about a median value of one. Data values ranged from zero to three with a mean of 0.882.

Table 6.9. Factor Level Data for Airspeed Error

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	N/A	0.70	0.57	1.38
Std Dev	N/A	0.82	0.73	0.81
Time of Day	*AM*		*PM*	
Average	0.62		1.15	
Std Dev	0.90		0.74	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	0.93	0.82	0.91	0.88
Std Dev	0.84	0.84	0.92	0.86
Task	*Primary*		*Primary with Secondary*	
Average	0.93		0.83	
Std Dev	0.96		0.76	

Results indicated average performance became progressively worse as Workload Level progressed from Monitor, through Nominal, to High Load levels. Workload Level resulted in an average performance error rating two times greater for the High Load treatment (1.37) than the medium treatments (0.64), $p < 0.001$.

In changing to the operationally related criterion of the Performance Rating there was one significant change in primary factors. Task, which was significant, $p < 0.001$, in the Composite Index became insignificant, but Time of Day became a significant factor, $F(1,12) = 82.05$, $p < 0.001$. All other interactions were the same.

Like the other performance indices for Workload, the Nominal treatments were not different but differed from the Monitor and High Load treatments, $p < 0.001$. There were 18 cases in which the rating was dangerous and an additional 193 segments in which subjects exceeded ATC criteria. The High Load treatment accounted for 15 of the 18 dangerous segments and 107 of the 193 segments exceeding ATC criteria. Trends in Performance Error Rating mirrored those shown in figures from the Composite Error.

Table 6.10. Five Factor ANOVA for Performance Error Rating

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.4701	0.4701	T x S(G)	2.67	NS
Group	[G]	3	0.2956	3.3385	S(G)	0.35	NS
#Subject(Group)	[S(G)]	12	13.3542	209.5000		3.06	
Time Day	[D]	1	13.2826	1.9427	D x S(G)	82.05	0.144
Workload Level	[W]	2	24.5000	2.0755	W x S(G)	141.65***	0.267
Interactions							
T x G		3	1.6706	2.1094	T x S(G)	3.17	NS
T x D		1	1.4180	1.0365	T x D x S(G)	16.42***	0.015
T x W		2	6.4245	2.6953	T x W x S(G)	28.60***	0.068
G x D		3	0.6706	1.9427	D x S(G)	1.38	NS
G x W		6	0.6536	2.0755	W x S(G)	1.26	NS
D x W		2	18.0339	3.4870	D x W x S(G)	62.06***	0.195
T x G x D		3	0.0456	1.0365	T x D x S(G)	0.18	NS
T x G x W		6	0.9427	2.6953	T x W x S(G)	1.40	NS
T x D x W		2	3.2578	1.0339	T x D x W x S(G)	37.81***	0.035
G x D x W		6	0.3958	3.4870	D x W x S(G)	0.45	NS
T x G x D x W		6	1.2083	1.0339	T x D x W x S(G)	4.68	0.010
Error Terms							
S(G)		12	3.3385	0.2782			
T x S(G)		12	2.1094	0.1758			
D x S(G)		12	1.9427	0.1619			
W x S(G)		24	2.0755	0.0864			
T x D x S(G)		12	1.0365	0.0864			
T x W x S(G)		24	2.6953	0.1123			
D x W x S(G)		24	3.4870	0.1453			
T x D x W x S(G)		24	1.0339	0.0430			
Total		191	90.9883				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

In the Time of Day factor, the afternoon treatment produced nearly twice the performance error of the morning treatments, $p < 0.001$ (0.85 vs 0.45). One hundred and twenty-eight small errors (one index point/exceeded ATC limits) occurred in the afternoon. Fifteen high error (two index points/dangerous condition) occurred in the morning. The Composite Error Index hid this qualitative difference in error type and frequency. Afternoon performance error for Nominal After High Load treatments following the High Load treatment level was greater than morning error for the same

conditions, $p < 0.001$, and greater than that of Nominal After Monitor treatments in the afternoon, $p < 0.001$. This result was consistent with previous performance results for the PM/Nominal After High Load treatment where a performance decrement was noted.

Eye movement results similar in nature to performance results are desirable from a modeling standpoint. To be similar, there must be predictable progression through the Workload Level (Monitor, Nominal, and High Load) treatments. Interactions similar to the Time of Day/Workload Level and Task/Time of Day/Workload Level interactions previously highlighted must be significant and correlated with the performance results. The High Load treatment and the Nominal After High Load interaction with Time of Day account for 100% of dangerous performance errors and 82% of all ATC deviations.

CHAPTER 7.

PSYCHOPHYSIOLOGICAL RESULTS

A total of 1024 data segments from 16 subjects were analyzed. Two samples, one with a primary task and one with both a primary and secondary task, were drawn from each of the 32 simulation segments flown. All subjects contributed 64 data segments. Twenty-six different eye movement parameters, two peripheral temperature parameters, and an EEG (Index of Engagement) parameter were associated with each segment.

Of the 1024 samples, eight lacked fixations despite valid eye tracking status. Since there were no fixations in 8 these segments, the 26 eye movement parameters were not available or were unreliable for those segments. Six of the 26 parameters were computed as the difference between two successive samples. If the comparison sample had no fixations, the parameter was not computed or used. This situation resulted in the loss of an additional seven data segments for change comparison variables using eye movements. Data from these additional seven samples were not included in the analysis.

Each treatment had four replicates per treatment for each subject during both morning and afternoon sessions. With the loss of the aforementioned samples, the number of replicates was reduced. The number of replicates for one subject was reduced to two in the case of one treatment. All other treatments had at least three replicates.

Several different elements reflected by psychophysiological parameters affect performance of any given task. Some elements described earlier in the Modified Human Information Processing (HIP) Model are attention (arousal), sensory processing (early perception), perception (perception strategy), and decision and response selection (cognitive processing) and will be considered below. In addition, psychophysiological

parameters showing a dependency on segment type will be discussed outside the context of the Modified HIP Model. These parameters will be discussed separately since they are more closely related to task than to the focus of this study, performance.

The first objective of this study was to measure aviation performance as related to normal workload, task overload, and task underload (Chapter 6). The second objective, linking workload to specific psychophysiological measures, is accomplished in this chapter. Finally, a relationship between eye movement parameters and performance will be described in Chapter 8.

7.1. Psychophysiological Parameters Related to Human Information Processing

When discussing the influence of various psychophysiological parameters it is convenient to associate them with different portions of the Modified Human Information Processing (HIP) Model previously presented. For example, pupil diameter is associated with level of arousal which relates to attention resources (Kahneman, 1973). Early Perception is reflected by basic eye movement data describing geometry of fixations as well as time and geometry of saccades. Perception Strategy is driven by the viewing strategies employed by subjects (where they look) as well as their visual acuity. (Visual acuity was controlled for this study.) Cognitive processing, which includes decision and response selection, has been tied to fixation duration (Just and Carpenter, 1976; Harris and Glover, 1985). Each of these areas would be expected to contribute to the response output from the model. Aviation performance is the response output in the case of the instrument cross check model.

7.2. Psychophysiological Variables

Table 7.1 summarizes the significance of the primary factors for 24 psychophysiological parameters. The assumption of equal variance was tested.

Table 7.1. Summary of Sample Variance and Factor Significance

Eye Movement Parameter/Factor	Secondary Task (T)	Group (G)	Subject (S[G])	Daytime (D)	Workload Level (W)
Arousal (Attention) Parameters					
Pupil Diameter	p<0.001	NS***	p<0.001***	NS**	p<0.001
Pupil Diameter Change	NS	NS**	NS***	NS	p<0.001
Peripheral Temp	NS	NS***	p<0.001***	p < 0.01**	p<0.001***
Change Peripheral Tmp	NS	NS***	NS***	p < 0.05**	p<0.001*
Early Perception (Sensory) Parameters					
Saccade Time	p<0.001***	NS***	p<0.001***	NS	p<0.001
Saccade Time Change	NS*	NS	NS***	p<0.05	p<0.01
Saccade Distance	NS*	NS***	p<0.001***	NS***	NS*
Saccade Dist Change	NS	NS	NS***	NS	NS
Fixation Size	NS*	NS***	p<0.001***	NS	NS
Fixation Size Change	NS**	NS*	NS***	p<0.01	p<0.01
Maximum Ellipticity	NS	NS***	p<0.001***	NS**	p<0.05*
Max. Ellipticity Change	NS	NS	NS***	NS	p<0.001
Strategy (Perception) Parameters					
Velocity Fix Gate	p<0.05	NS***	p<0.001*	p < 0.05***	p<0.01*
Angle Fix Gate	NS	NS***	p<0.001***	p < 0.05*	p<0.05
Dual Fixation Gate	p<0.001***	NS**	p<0.001***	NS	p<0.001
Trans Matrix Symmetry	NS	NS***	p<0.001***	NS	NS
Trans Matrix Repeat	NS	NS***	p<0.001***	NS	p<0.001
Trans Matrix Useful	NS	NS***	p<0.001***	NS***	p<0.001
Short Fixations	NS	NS	p<0.001***	NS	NS
Number of Cycles	NS	NS***	p<0.001***	NS**	p<0.001*
Cognitive Processing (Decision and Response) Parameters					
Fixation Time	p<0.05*	NS***	p<0.001***	NS	p<0.001
Fixation Time Change	p<0.05	p<0.05	NS***	NS	p<0.001
Long Fixations	p<0.05***	NS***	p<0.001***	NS	p<0.001
Index of Engagement	p<0.001	NS	p<0.001***	NS	p<0.001

Note: NS – Not Significant

Note: Significance of Heteroscedastcity, * - p < 0.05, ** - p < 0.01, *** - p < 0.001

Heteroscedasticity was significant among Subject(Group) and Group for all parameters. The level of heteroscedasticity is indicated by the asterisks accompanying the significance level. Many of the parameters were significant despite heteroscedasticity. Parameters with significant factors which displayed heteroscedasticity were analyzed using a Greenhouse-Geisser (1959) model to account for the heteroscedasticity.

Significance in the Workload Level factor is of particular interest because of its relationship to performance. Since performance error increased across Workload Levels, significant differences in psychophysiological parameters across Workload Levels indicate a potential relationship with performance. The relationship between performance and Workload Level was detailed in Chapter 6.

7.3 Presentation Method for Eye Movement Results

Nineteen eye movement parameters demonstrated significant differences among Workload Levels. Of these nineteen parameters, five did not parallel changes in performance across the Workload Levels. These five parameters will be presented and compared to the other psychophysiological data but not workload conditions.

The number of cycles was significantly different among workload levels, however the four Cycle parameters related to viewing patterns will not be developed due to the large number of treatments with no cycles. Discussion of the cycle parameters will elaborate on this topic.

The results will be presented in the order in which parameters appear in Table 7.1. Of the multiple factor interactions, two, 2-factor interactions were common. The Task/Workload interaction will be presented in a dashed line format. The Time of Day/Workload interaction will be presented in a solid line format to differentiate it from Task/Workload Level results. The three factor interaction, Task/Time of Day/Workload Level will be presented in a four panel line graph format. This interaction is important to understanding an afternoon performance decrement.

7.4. Psychophysiological Parameters Related to Arousal

Pupil diameter and changes in pupil diameter are associated with arousal (Kahneman, 1973). In addition, this study demonstrated the correlation of peripheral temperature variables with workload. However, the peripheral temperature variables did not display interactions corresponding to the interactions from performance and other psychophysiological parameters. The peripheral temperature results were not reactive to unique changes in workload due to secondary task. Nor did they display the interaction associated with the PM/Nominal After High Load treatment. Peripheral temperature results were similar to pupil diameter in this regard. Thus, peripheral temperature was grouped with arousal (attention) parameters where its relationship with "fight or flight" mechanisms fits naturally.

Measuring the change in pupil diameter and peripheral temperature was an effective method to reduce the variability of the parameters. Pupil Diameter Change and Peripheral Temperature Change both had reduced heteroscedasticity when compared to the data of the basic measurements of Pupil Diameter and Peripheral Temperature. The

reductions in variance were accompanied by large increases in the amount of variance accounted for by significant factors of the ANOVA. Peripheral Temperature accounted for 65.9% of ANOVA variance, making it a good candidate for modeling performance.

7.4.1. Pupil Diameter

Pupil diameter is one of the eleven eye movement parameters with a discernable relationship to performance across Workload Levels. Pupil Diameter was a direct output of the eye tracker measured in pixels at the digital interface of the infrared camera. It is an average value from all eye tracker samples used to calculate fixation parameters in the given data segment.

Average Pupil Diameter was normally distributed and ranged from 64 to 246 with an average of 112.66. The standard deviation was 25 pixels and was homogenous for significant factors. The average pupil diameter Five-Factor ANOVA for Task and Workload Level may be found in Appendix E8. Treatments within both Task $F(1,12) = 22.19$, $p < 0.001$, and Workload Level, $F(3,36) = 11.22$, $p < 0.001$, were significantly different. The interaction of the two factors was also significant $F(3, 36) = 6.18$, $p > 0.01$.

Task accounted for more variance than any other factor. Average values for pupil diameter are shown in Table 7.2. Subjects were consistently more wide eyed (aroused) with the secondary task. The spread of seven pixels between factors was greater than the spread across Workload Level. The spread across Subject Groups was 20 pixels.

Table 7.2. Factor Level Data – Pupil Diameter (horizontal pixels)

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	109	114	113	115
Std Dev	24	25	24	23
Time of Day	***AM***		***PM***	
Average	114		111	
Std Dev	27		21	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	123	103	106	117
Std Dev	21	25	14	27
Task	***Primary***		***Primary with Secondary***	
Average	109		116	
Std Dev	22.6		24.6	

Differences among the Workload Level treatments were significant, but not as significant as with Performance Error Rating. The relationship among treatments of Performance Rating (Fig. 7.1) and Average Pupil Diameter (Fig. 7.2) were remarkably similar in form. Medium treatments of both dependent variables were not significantly different. The Monitor treatment of Pupil Diameter was different from the High Load treatment, $p < 0.05$; that was the only significant difference among Workload Level treatments due to high variance. For Performance Error Rating both Monitor and High Load treatments were significantly different from the nominal levels as well as from each other, $p < 0.001$.

Like many variables, Average Pupil Diameter increased in a consistent manner with the addition of the secondary task. The Task/Workload Level interaction resulted from treatments with the secondary task being similar, while Primary Task/High Load treatments were different, $p < 0.001$.

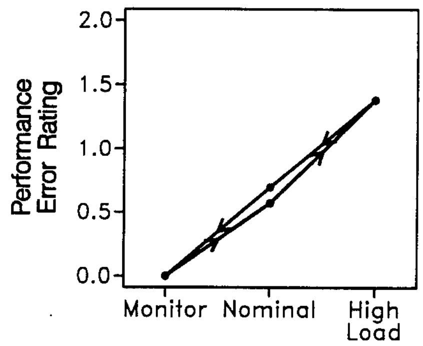

Figure 7.1. Performance Rating vs. Workload Level

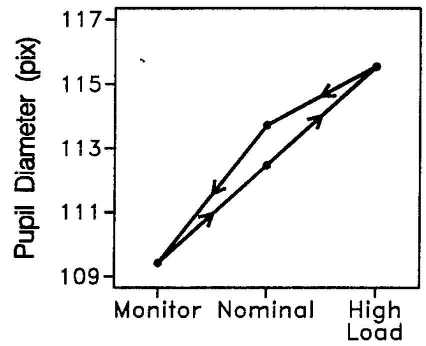

Figure 7.2. Average Pupil Diameter vs. Workload Level

Performance error increased in the Nominal After High Load treatment with Primary Task only. However, in the case of Pupil Diameter, the same increase was associated with the Primary and Secondary Task treatments. Pupil Diameter did not correspond to performance decrements for Nominal treatments but did correspond at the High Load treatment. This difference indicates the underlying mechanism was related to the performance error for the High Load treatment (task overload) was different from that correlated to the performance error for the Nominal After High Load treatment.

Average Pupil Diameter was one of the few parameters without a significant Secondary Task/Time of Day/Workload Level interaction.

7.4.2. Pupil Diameter Change

The range of pupil diameters varied for each subjects. The average change in pupil diameter was used as a basis of comparison to remove subject differences in pupil diameter. Pupil Diameter Change was determined by subtracting the value of the previous segment Average Pupil Diameter from that of the most recent segment. Since

each segment possessed two data samples (one with a secondary task present and one without the task), the sample with the similar task was used for comparison.

Pupil Diameter Change varied from –55.7 pixels to 51.9 pixels. The average was 0.3 pixels with a standard deviation of 10 pixels. The data was normally distributed around a mean of 0.431 pixels. The factors and interactions of Task, Time of Day, and Workload Level displayed uniform variance.

Table 7.3. Factor Level Data – Pupil Diameter Change

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	-3.52	-1.19	4.26	3.67
Std Dev	14	14	13	13
Time of Day	***AM***		***PM***	
Average	0.55		1.08	
Std Dev	14		14	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	1.09	1.32	0.65	0.19
Std Dev	13	20	11	10
Task	***Primary***		***Primary with Secondary***	
Average	0.73		0.89	
Std Dev	12		15	

Results of the five factor ANOVA for Pupil Diameter Change may be found in Table E9 of Appendix E. One factor, Workload Level, was significant, $F(3, 36) = 14.99$, $p < 0.001$. The significant interactions were between Task/Workload Level, $F(3, 36) = 4.38$, $p < 0.01$, and among Task/ Time of Day/Workload Level, $F(3, 36) = 4.30$, $p < 0.05$. Significant factors and interactions accounted for 25.2% of the ANOVA variation.

Eight of 1024 segments had invalid data affecting the calculation of this parameter for 15 data segments. The total number of data samples used was 1009.

A majority of the variation in Pupil Diameter Change was accounted for by Workload Level. Figure 7.3 illustrates the increase of Average Change in Pupil Diameter

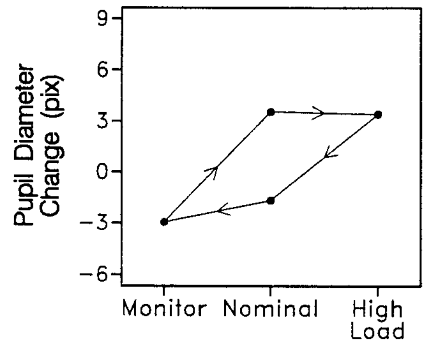

Figure 7.3. Pupil Diameter Change versus Workload

which was similar to increases in performance error. When the task became easier (Monitor or Nominal after High Load treatments), there was no significant difference between the treatments. Nor was there a significant difference when the workload increased (High Load or Nominal After Monitor treatments). However, there was significant difference between factors with increasing workload versus decreasing workload, $p < 0.001$.

Figure 7.4 displays the effect of the Task and Workload Level interacting. Addition of the Secondary Task resulted in a more regular, predictable progression in the variable. No adjacent Workload Levels had significantly different results with the Primary with Secondary Task treatment. Without the Secondary Task, results were polarized according to whether there was an increase or decrease in workload, $p < 0.001$. Examples of the Task/Workload two factor interaction are shown with dashed lines to

avoid confusion with examples of the two factor Time of Day/Workload interaction which are shown with solid lines.

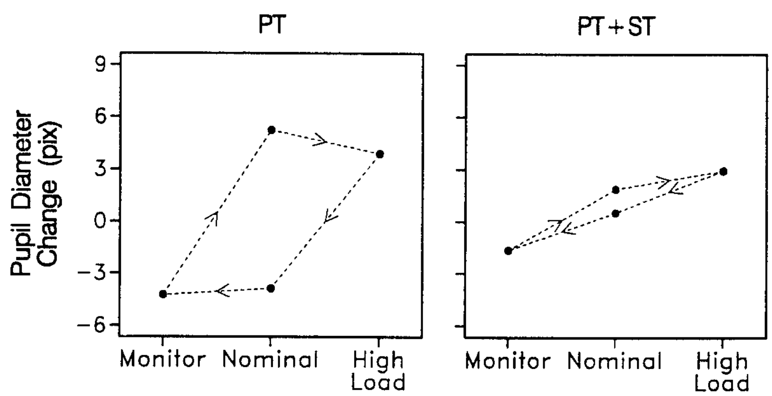

Figure 7.4. Pupil Diameter Change interaction with Task/Workload

The three factor interaction of Task/Time of Day/Workload Level (Fig 7.5) occurred because all factor levels of workload were the same for the Primary Task/AM treatments. Morning results, with the secondary task, were equivalent ($p>0.999$). There was not a change in the level of arousal when workload was changed. All other treatment combinations showed a significant difference between Nominal treatments ($p<0.05$).

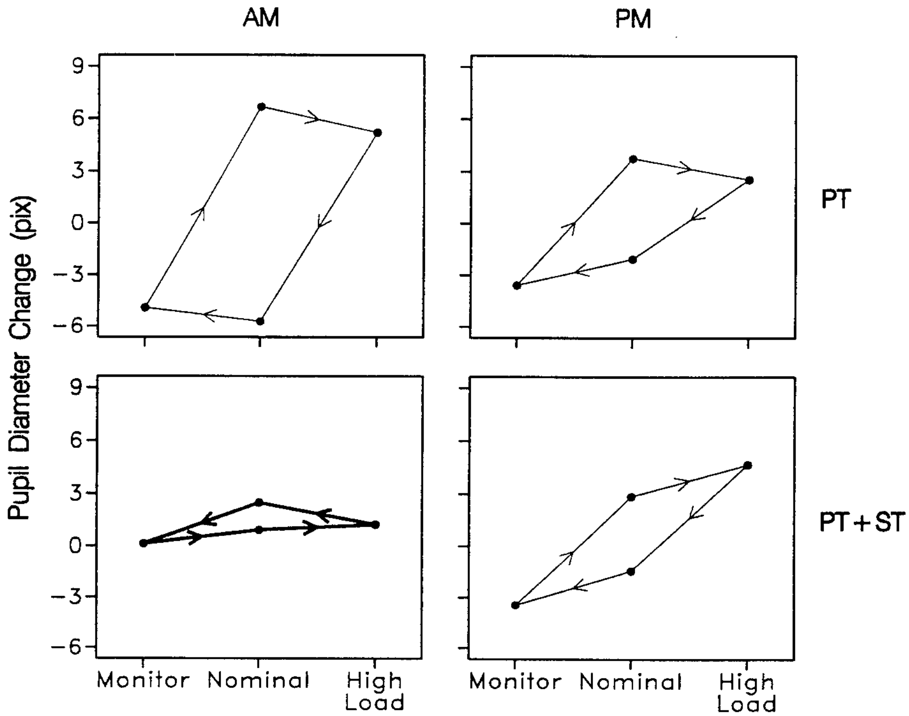

Figure 7.5. Pupil Diameter Change interaction with Task/Time of Day/Workload

7.4.3. Peripheral Temperature

Peripheral temperature was measured continuously by a sensor attached with Velcro to the proximal portion of the left index finger on the dorsal surface of the hand. The reading was recorded in two different forms. It was digitally recorded at 2 Hz and it was recorded on a strip chart video output. Since some of the digital peripheral temperature data was corrupted, data was read off of the video recording for all subjects.

The mean peripheral temperature was 81.09° F and range was 71-92° F. Peripheral temperature data was normally distributed with a slight break at the median value of 81°F which makes the distribution appear slightly bimodal. Standard deviation was 5.4° F. Workload Level, Time of Day, and Group factors showed significant heteroscedasticity.

The factors of Subject, Time of Day, and Workload Level were significant along with five interactions. The significant factors and interaction accounted for 20.6% of ANOVA variance.

Subjects had significantly higher peripheral temperature in the afternoon than in the morning, $F(1, 12) = 12.31, p < 0.01$. Table 7.4 shows a three degree increase between Time of Day mean temperatures. This increase indicated the subjects were less aroused during the afternoon simulation. A Time of Day/Group interaction occurred because the average peripheral temperature for the NASA technicians did not change for the afternoon session while the other groups increased peripheral temperature by at least three degrees.

Table 7.4. Factor Level Data – Peripheral Temperature

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	82.71	80.04	82.13	79.49
Std Dev	5.80	4.96	5.37	4.60
Time of Day	***AM***		***PM***	
Average	79.65		82.53	
Std Dev	5.43		4.91	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	81.22	81.33	83.85	77.99
Std Dev	5.62	5.90	4.51	4.71
Task	***Primary***		***Primary with Secondary***	
Average	79.65		82.53	
Std Dev	5.43		5.42	

Workload Level treatments were polarized in two like groups. The High Load and Nominal After High Load treatments composed one group, and the Monitor and Nominal After Monitor treatments composed a second group. This pairing differed for

pupil diameter. The two groups differed, $F(3, 36) = 55.20$, $p < 0.001$, by over two degrees. Although the average temperatures changed slightly in the Nominal treatments, they tended to take on the temperature profile of the segment followed. The Peripheral Temperature Change results and discussion will further explicate the tendency to mimic the previous segment.

The lower temperatures found in the High Load and Nominal After High Load treatments correspond to increases in performance error occurring in the same treatments. If peripheral temperature was related to error induced stress, this relationship should become more evident within the context of the Workload Level/Time of Day interaction where the Nominal After High Load/PM treatments were a unique source of performance error.

As expected, two different workload profiles were associated with the two Time of Day factor levels creating a significant interaction between Time of Day and Workload, $F(3, 36) = 42.05$, $p < 0.001$. In the morning, Nominal conditions were alike and both Nominal/AM treatments were significantly different from the other morning workload treatment. In the morning, peripheral temperature changed with workload.

However, in the afternoon nominal treatments were different from each other, $p < 0.001$. In fact, the Nominal treatments did not change from the High Load or Monitor treatment which they followed. In the morning, peripheral temperature reacted to workload changes, but in the afternoon some inertia had developed in the reaction mechanism. This anomaly will be evident in the Peripheral Temperature Change results.

The High Load/PM, Nominal After High Load/PM and High Load/AM treatments were the three greatest sources of performance error. These three treatments

also reflect the lowest peripheral temperatures for their respective times of day. The Nominal After High Load/PM treatment which was the source of the Workload Level/Time of Day interaction for performance error was also the source of the interaction for Peripheral Temperature.

Finally, the Time of Day/Workload Level interaction helped determine the source of peripheral temperature change. If temperature variation was due to changing stress levels, it was possible that the Nominal After Monitor and Nominal After High Load tasks would have different results despite being the same physical task. This was the case for the afternoon treatments, demonstrating the link between peripheral temperature and stress.

The three factor interaction, Workload Level/Task/Time of Day, was a result of the large increase in temperature for the Nominal After Monitor/P or P+S Task/PM treatments, $p < 0.001$. This result was consistent with performance results performance results. Figures included in Peripheral Temperature Change will aid in explaining this interaction.

7.4.4. Peripheral Temperature Change

Peripheral Temperature Change was computed as the difference between the current segment peripheral temperature and the peripheral temperature from the previous segment of the instrument pattern. This differential was selected so the Nominal Workload Level treatments would be compared against the Monitor or Workload segment it followed. In the morning, the Nominal After Monitor and Nominal After High Load treatments both occurred on downwind segments. In the afternoon both Nominal treatments occurred on

approach segments. Thus, the physical control task was the same for both Nominal treatments for the given time of day.

Peripheral Temperature Change data ranged from –10°F to 10°F with a mean of –0.00981°F. Data was normally distributed around a median value of zero. Standard deviation was 2.99°F. Data for factor levels within Group, Time of Day, and Workload Level displayed heteroscedasticity. All significant p-values for heteroscedastic data were verified using Greenhouse-Geisser (1959) methodology.

Results of the five factor ANOVA for Peripheral Temperature Change are in Table 7.5. Time of Day, and Workload Level were significant factors and six interactions were significant. Significant factors and interactions accounted for 66.7% of ANOVA variance.

Peripheral Temperature Change increased 0.1°F for the afternoon ($\omega^2 = 0.2\%$, $F(1, 12) = 6.30$, $p < 0.05$), and there was a corresponding increase in Performance Error Rating for the afternoon treatments. The baseline temperature increased slightly for the afternoon session which was expected as a result of natural relaxation after lunch. The Time of Day/Workload Level interaction for Peripheral Temperature Change also indicated a shift in arousal for some of these treatments.

Workload was a significant factor, $\omega^2 = 46\%$, $F(3, 36) = 38.54$, $p < 0.001$. Nominal treatments within the Workload Level factor were significantly different, $p < 0.05$.

Table 7.5. Five Factor ANOVA for Peripheral Temperature Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.0885	0.0885	T x S(G)	2.07	NS
Group	[G]	3	0.3061	0.1020	S(G)	0.14	NS
#Subject(Group)	[S(G)]	12	34.1114	2.8426		0.55	
Time Day	[D]	1	2.6596	2.6596	D x S(G)	6.30*	0.002
Workload Level	[W]	3	613.4975	204.4992	W x S(G)	38.54***	0.464
Interactions							
T x G		3	0.0084	0.0028	T x S(G)	0.07	NS
T x D		1	0.2048	0.2048	T x D x S(G)	5.29*	0.000
T x W		3	20.4355	6.8118	T x W x S(G)	7.09***	0.014
G x D		3	1.1551	0.3850	D x S(G)	0.91	NS
G x W		9	112.0850	12.4539	W x S(G)	2.35*	0.050
D x W		3	119.2172	39.7391	D x W x S(G)	24.90***	0.089
T x G x D		3	0.0953	0.0318	T x D x S(G)	0.82	NS
T x G x W		9	10.7726	1.1970	T x W x S(G)	1.24	NS
T x D x W		3	18.5788	6.1929	T x D x W x S(G)	8.77***	0.013
G x D x W		9	53.0264	5.8918	D x W x S(G)	3.69**	0.030
T x G x D x W		9	6.9111	0.7679	T x D x W x S(G)	1.09	NS
Error Terms							
S(G)		12	8.5278	0.7107			
T x S(G)		12	0.5131	0.0428			
D x S(G)		12	5.0683	0.4224			
W x S(G)		36	190.9999	5.3056			
T x D x S(G)		12	0.4644	0.0387			
T x W x S(G)		36	34.6117	0.9614			
D x W x S(G)		36	57.4640	1.5962			
T x D x W x S(G)		36	25.4329	0.7065			
Total		255	1282.1241				

* = $p<0.05$ ** = $p<0.01$ *** = $p<0.001$ NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table 7.6. Factor Level Data – Peripheral Temperature Change

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	2.25	0.37	-0.57	-2.05
Std Dev	2.78	2.35	2.59	2.49
Time of Day	*AM*		*PM*	
Average	-0.11		0.10	
Std Dev	2.74		3.22	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	-0.02	-0.04	-0.01	0.06
Std Dev	2.26	3.22	3.57	2.77
Task	*Primary*		*Primary with Secondary*	
Average	-0.03		0.02	
Std Dev	3.01		2.97	

In addition, Figure 7.6 shows all other Workload Level treatments were different from each other, $p < 0.001$. By comparison, Nominal treatments for performance error were not significantly different, otherwise Peripheral Temperature Results show a consistent negative correlation to performance error. Peripheral Temperature Change decreases as performance error increases.

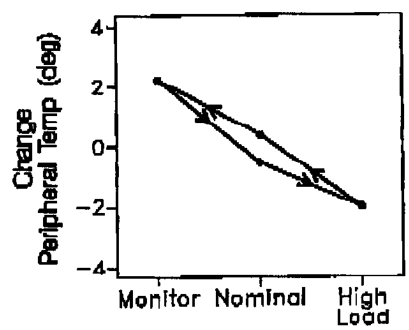

Figure 7.6. Peripheral Temperature Change versus Workload Level

The Group/Workload Level interaction was shown (Fig. 7.7) to illustrate the different strategy employed by commercial pilots. On the Nominal (Medium) treatments the commercial pilots showed more concern when workload was reduced for the Nominal After High Load segments, than when it was increased for the Nominal After Monitor segments. Commercial Pilots, as a group, had a more consistent and lower peripheral temperature profile. The format shown in Figure 7.7 was chosen to illustrate the difference for commercial pilots who had experience in the study environment and an aviation rating. During debrief, the commercial pilot group was the only group stating an expectation that their response to emergencies would be tested during the simulation.

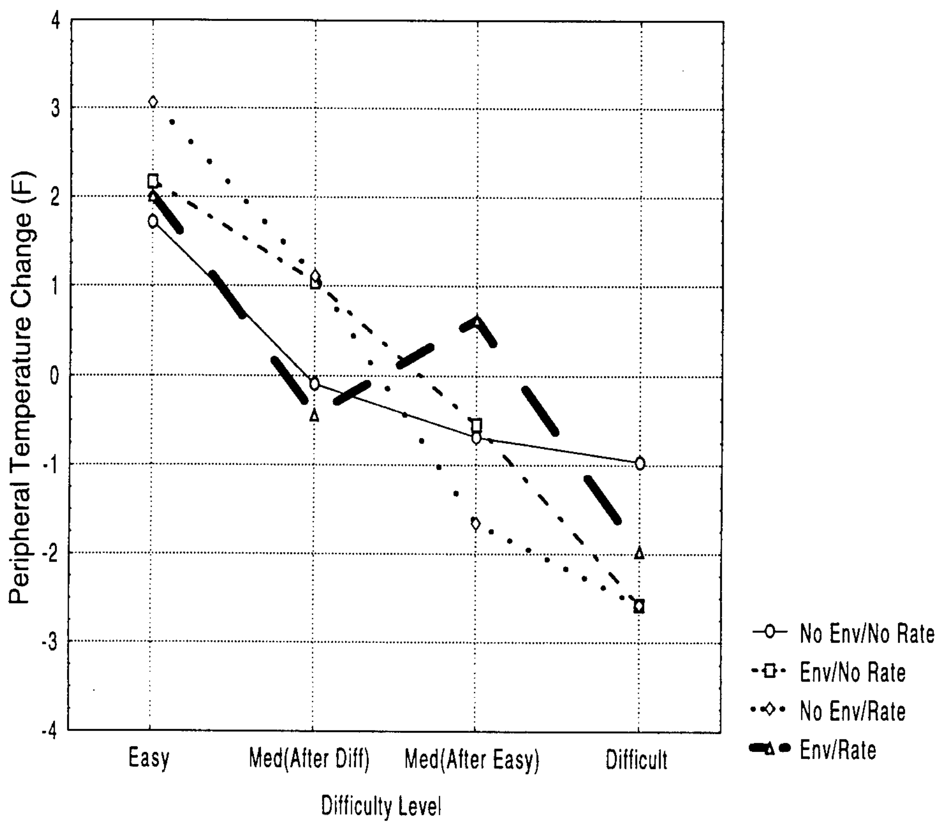

Figure 7.7. Peripheral Temp Change interaction with Workload Level/Group

Results from many variables demonstrated a Workload Level/Time of Day interaction occurring as a result of the Nominal After High Load treatment. Peripheral Temperature Change showed this interaction as well. The Nominal After Monitor treatments were different between morning and afternoon (Figure 7.8). However, it was the afternoon treatment with virtually no change that was surprising. Subjects did not change their level of arousal with change in workload for this treatment.

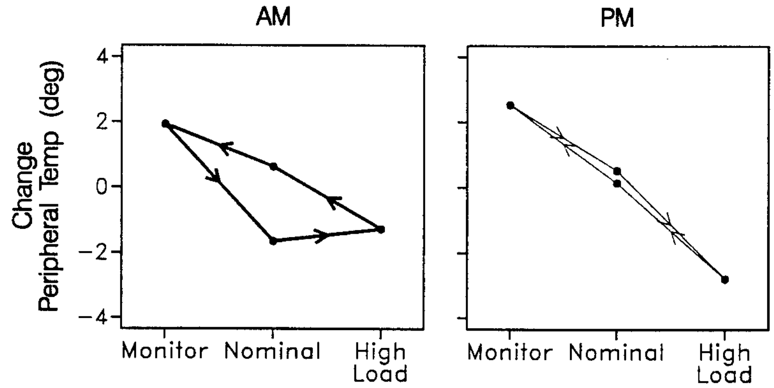

Figure 7.8. Peripheral Temp Change interaction with Workload Level/Time of Day

The three factor interaction in Figure 7.9 bears out the similarity of nominal conditions for the afternoon treatments. The PM/Primary Task treatment highlights the lack of peripheral temperature change for the nominal workload treatments. This single treatment was also responsible for significant performance error. In addition, the PM/Primary with Secondary Task treatment also showed little change in peripheral temperature for the nominal factor levels. Differences between morning and afternoon results, particularly PM/Primary Task treatments, $p < 0.001$, are graphically illustrated by comparison of the left and right panels of Figure 7.9.

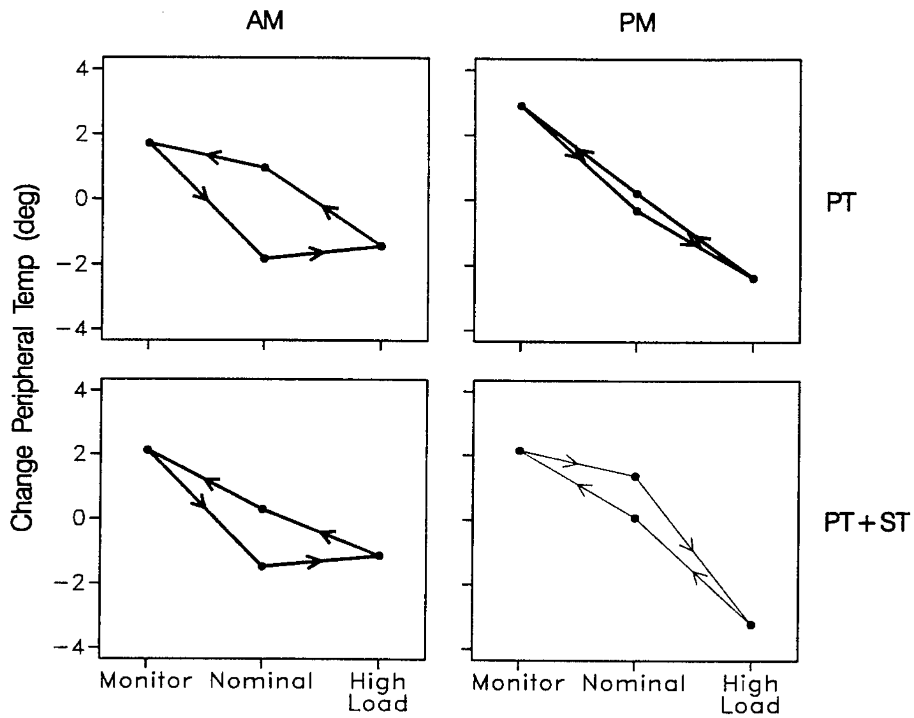

Figure 7.9. Peripheral Temp interaction with Diff Lvl/2nd Task/Time of Day

7.5. Early Perception Variables

The Modified HIP Model has a stage labeled "Sensory Processing". To avoid confusion with discussion of cognitive processing this stage is referred to as "Early Perception". Psychophysiological parameters related to this stage are those describing the basic eye movement characteristics that are independent of scan patterns and cognitive processing. Sensory Function parameters describe how subjects look at something, not where they look or how they process the information gleaned. Saccadic measures are included in this section since little or no cognitive processing takes place during saccadic movement (Biederman, 1991).

A total of eight parameters, four of the basic parameters and their accompanying "Change" parameters, were considered Sensory Parameters. They include Saccade Time, Saccade Distance, Fixation Size, and Maximum Ellipticity. Four of the eight, Fixation Size, Fixation Size Change, Maximum Ellipticity, and Maximum Ellipticity Change were task dependent. These parameters correlated to the subtask level. It was not the intent of this study to differentiate among segment types. Rather, it was the purpose of this study to differentiate among workloads. Therefore, these parameters are not considered any further for the purposes of this study.

Saccade Distance and Saccade Distance Change did not correlate to any discernable pattern seen among the other parameters or factors. Distances were based on calculations made in the analysis code which translated fixation positions in different oculometer scene planes. This was a two step process which first determined fixation position within the scene plane and then accounted for geometry between scene planes. Although data quality indicates fixations mapped correctly into areas of interest within the scene planes, it is possible the geometry among scene planes was not correctly calculated to translate the eight scene planes into one frame of reference. Further analysis of scene plane geometry is required to determine the importance of saccadic distance.

7.5.1 Saccade Time

Saccade time was measured from the last point falling within the fixation to the first point falling within the subsequent fixation. The average saccade time for 1016 segments ranged from 25 to 57 milliseconds (ms) with a mean of 42 ms. Data was normally distributed about a median of 42 ms with a standard deviation of five

milliseconds. The factors of Time of Day and Workload Level were significant but only Workload Level displayed uniform variance. One interaction, Time of Day/Workload Level was significant.

Results of the five factor ANOVA are found in Table 7.7. Average Saccade Time increased systematically with Workload Level, $F(3, 36) = 13.23$, $p < 0.001$ in a manner comparable to Performance Error Rating. Task was also significant , $F(1,12) = 41.36$, $p < 0.001$. One significant interaction occurred among Task/Group/Time of Day,

Table 7.7. Five Factor ANOVA for Saccade Time (milliseconds)

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.000593	0.000593	T x S(G)	42.35***	0.128
Group	[G]	3	0.000426	0.000142	S(G)	0.94	NS
#Subject(Group)	[S(G)]	12	0.007264	0.000605		68.13***	
Time Day	[D]	1	0.000020	0.000020	D x S(G)	0.63	NS
Workload Level	[W]	3	0.000156	0.000052	W x S(G)	13.23***	0.032
Interactions							
T x G		3	0.000068	0.000023	T x S(G)	1.63	NS
T x D		1	0.000000	0.000000	T x D x S(G)	0.00	NS
T x W		3	0.000016	0.000005	T x W x S(G)	1.93	NS
G x D		3	0.000061	0.000020	D x S(G)	0.63	NS
G x W		9	0.000028	0.000003	W x S(G)	0.80	NS
D x W		3	0.000068	0.000023	D x W x S(G)	7.47***	0.013
T x G x D		3	0.000042	0.000014	T x D x S(G)	2.08	NS
T x G x W		9	0.000042	0.000005	T x W x S(G)	1.66	NS
T x D x W		3	0.000019	0.000006	T x D x W x S(G)	2.21	NS
G x D x W		9	0.000026	0.000003	D x W x S(G)	0.97	NS
T x G x D x W		9	0.000019	0.000002	T x D x W x S(G)	0.76	NS
Error Terms							
S(G)		12	0.001816	0.000151			
T x S(G)		12	0.000168	0.000014			
D x S(G)		12	0.000389	0.000032			
W x S(G)		36	0.000142	0.000004			
T x D x S(G)		12	0.000080	0.000007			
T x W x S(G)		36	0.000101	0.000003			
D x W x S(G)		36	0.000109	0.000003			
T x D x W x S(G)		36	0.000102	0.000003			
Total		255	0.004493				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

$F(3,12) = 1.81$, $p < 0.05$. Novice group saccade time was the same for both tasks in the morning. All other Group comparisons between morning and afternoon were different, $p<0.001$. Significant ANOVA factors accounted for 17.4% of the variance.

The Task factor accounted for 13% of the ANOVA variance. The Primary Task treatments were an average of 2 ms (5%) longer than the Primary with Secondary Task treatments. Performance was also better for Primary with Secondary Task treatments pointing toward a correlation between shorter saccade times and better performance for aviation tasks.

Table 7.8. Factor Level Data – Saccade Time

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.041	0.042	0.042	0.043
Std Dev	0.005	0.005	0.005	0.005
Time of Day	***AM***		***PM***	
Average	0.042		0.042	
Std Dev	0.005		0.005	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.040	0.042	0.043	0.043
Std Dev	0.005	0.005	0.004	0.006
Task	***Primary***		***Primary with Secondary***	
Average	0.043		0.043	
Std Dev	0.005		0.004	

Figure 7.10 illustrates the systematic increase of saccade time with level of Workload. Like the performance variables, there was no significant difference between the two Nominal conditions. The High Load treatment was significantly different from the Monitor treatment, $p < 0.001$, and the Nominal treatments, $p < 0.05$. Like the performance variables, the High Load and Monitor treatments were different from the

Nominal treatments, $p < 0.01$. Nominal treatments appear to be similar, however the reason for this misleading appearance will be explained by interaction results.

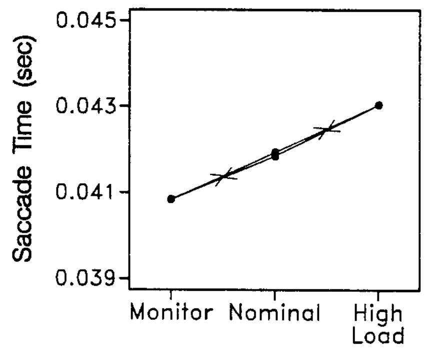

Figure 7.10. Average Saccade Time vs. Workload Level

The Saccade Time interaction of Time of Day /Workload Level again provides evidence of a shift in the subjects approach to the two Nominal treatments between the morning and afternoon simulations. Figure 7.11 illustrates the variables similarity to performance error. Treatment data was very reactive to changes in Workload Level in the morning but not so much in the afternoon. The Monitor and Nominal After Monitor treatments both displayed similar Average Saccade Times. These two groupings are comparable to groupings for performance error. Both Nominal treatments in the afternoon were significantly different from their morning counterparts, $p < 0.001$. The Afternoon/Nominal After High Load treatment was significantly greater than it's morning counterpart while the Afternoon/Nominal After Monitor treatment was significantly lower, $p = 0.001$. Thus the reason for the clockwise morning pattern versus the counter-clockwise afternoon pattern. This was also the reason the two nominal treatments displayed no difference for the Workload factor.

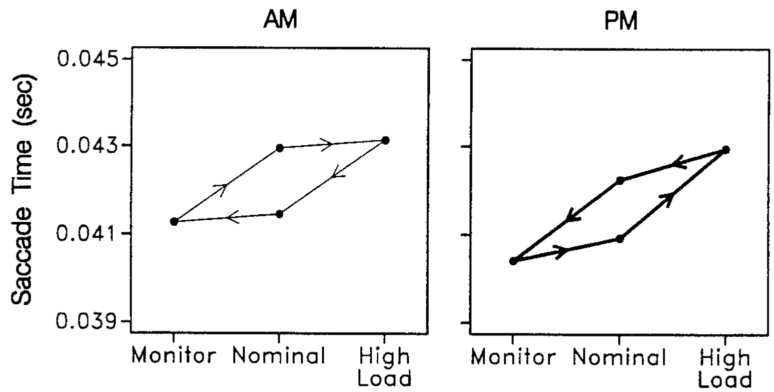

Figure 7.11. Average Saccade Time interaction of Time of Day/Workload Level

7.5.2. Saccade Time Change

Saccade time was measured from the last point included in a fixation to the first point included in the subsequent fixation. Multiple saccades could take place between fixations. Movement from one fixation to the next often involves corrective saccades when the target is not acquired on the first move (Boff and Thomas, 1996). Change in saccade time was determined by comparison of the current data segment average saccade time to the average saccade time of the previous segment with like Task. Values were normally distributed between –43 and 16 milliseconds with a median value of –0.03 sec and standard deviation of 5 sec.

Appendix E, Table E13, shows the results for the five factor ANOVA. Workload Level, $F(3,12) = 6.60$, $p < 0.01$, and Time of Day, $F(1,12) = 8.20$, $p < 0.05$, produced significant results. Significant factors accounted for 9.7% of ANOVA variance. The 33 millisecond resolution available with a 60 Hz. Eye tracker limited resolution of factors and interactions. This was worst case resolution resulting when a saccade was initiated at

the instant of the last oculometer measurement within a fixation and ended at the instant a new fixation began.

The Time of Day data indicates an increase of saccade time through the morning simulation but a decrease in the afternoon simulation. No comparable change was present for performance. However, this parameter was another indicator of a change in perception strategy between morning and afternoon. There was a significant Task/Time of Day/Workload interaction involving the Primary/PM/Nominal After High Load treatment. This was the same interaction highlighted in performance results.

Table 7.9. Factor Level Data – Saccade Time Change

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	-0.000746	-0.001044	0.001184	0.000444
Std Dev	0.04280	0.004410	0.004392	0.004407
Time of Day	*AM*		*PM*	
Average	0.000090		-0.000171	
Std Dev	0.004367		0.004548	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	-0.000043	0.000142	-0.000181	-0.000076
Std Dev	0.004079	0.004792	0.004265	0.004682
Task	*Primary*		*Primary with Secondary*	
Average	0.000018		-0.000099	
Std Dev	0.004671		0.004237	

7.5.3. Saccade Distance

Saccade distance was measured from the last point included in the previous fixation to the first point of the subsequent fixation. The distance was not an oculometer output but was calculated from the scene plane information associated with each fixation. Data was distributed as an exponential decay. There was one significant interaction

involving Time of Day/Workload Level. However, there were no similarities to performance data or any other results in this study.

7.5.4. Saccade Distance Change

The Saccade Distance change was computed as the difference in Average Saccade Time from the current segment to the previous segment with like Secondary Task treatment. Data showed no identifiable distribution trends. There was no significant factor and the only significant interaction, Time of Day/Workload, bore no resemblance to other data presented in this study.

7.5.5. Average Fixation Size

Average Fixation Size was the average distance measured from the center of the fixation to the points included in the fixation. Average Fixation Size ranged from 0.021 to 1.13 inches with an average size of 0.27 inches. Standard deviation was 0.12 inches for the normally distributed data. The only significant factor was subject. However, the three interactions commonly significant in this study were significant here, too. Task/Workload Level, $F(3,36) = 8.19$, $p < 0.001$, Time of Day/Workload Level, $F(3,36) = 13.31$, $p < 0.001$, and Task/Time of Day/Workload Level, $F(3,36) = 4.86$, $p < 0.01$ accounted for 6.3% of the ANOVA variation. These interactions were similar to those described for Saccade Time, but the significance level and variation accounted for were not as great.

Table 7.10. Factor Level Data – Average Fixation Size

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	0.294	0.276	0.282	0.291
Std Dev	0.121	0.116	0.118	0.108
Time of Day	*AM*		*PM*	
Average	0.274		0.298	
Std Dev	0.121		0.110	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	0.254	0.309	0.279	0.301
Std Dev	0.138	0.121	0.079	0.111
Task	*Primary*		*Primary with Secondary*	
Average	0.272		0.300	
Std Dev	0.104		0.125	

7.5.6. Fixation Size Change

Fixation Size Change was calculated as the difference between the current segment average Fixation Size and the average Fixation Size from the previous segment with a like Task treatment. Data was normally distributed with a mean of –0.0017 inches. The data range was from –0.540 to 0.450 with a standard deviation of 0.11 inches. The data displayed homogenous variance for the significant factors of Time of Day, $F(1,12) = 10.79$, $p < 0.001$, and Workload Level, $F(3,36) = 4.20$, $p < 0.05$. The interactions of Task/Workload Level, $F(3,36) = 4.86$, $p < 0.01$, and Time of Day/Workload Level, $F(3,36) = 17.61$, $p < 0.001$, were significant. The significant factors and interactions accounted for 11.1 % of the ANOVA variance.

The Time of Day Workload Level/ Workload Level interaction with Fixation Size Change was like the same interaction shown for Velocity Fixation Gate and Maximum Ellipticity. The interaction appears related to the nature of the task performed. The approach segments account for the highest values at each Workload Level.

Table 7.11. Factor Level Data – Fixation Size Change

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	0.0143	-0.0218	-0.0140	0.0147
Std Dev	0.110	0.115	0.119	0.110
Time of Day	*AM*		*PM*	
Average	-0.005064		0.001655	
Std Dev	0.107		0.122	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	-0.003725	-0.002837	-0.000050	-0.000266
Std Dev	0.125	0.123	0.101	0.107
Task	*Primary*		*Primary with Secondary*	
Average	-0.003524		0.000098	
Std Dev	0.104		0.124	

7.5.7. Maximum Ellipticity

To determine if changes in fixation geometry occur with factor levels, the average ellipticity of fixations was recorded for each segment. Maximum ellipticity was the distance to the farthest point within the fixation measured from the average fixation diameter. Maximum ellipticity ranged from 0.05 to 1.53 inches with a variance of 0.05 inches^2. This parameter was normally distributed around a median on 0.77 inches but heteroscedasticity was significant, $p < 0.05$.

The factors of Subject and Workload Level were significant. The significant interactions included Task/Workload Level, $F(3,36) = 10.60$, $p < 0.001$, Time of Day/Workload Level, $F(3,36) = 15.21$, $p < 0.001$, and Task/Time of Day/Workload Level, $F(3,36) = 5.62$, $p < 0.01$. ANOVA significant factors and interactions accounted for 12.6% of the variance. The complete ANOVA table may be found in Appendix E, Table E8.

Table 7.12. Factor Level Data – Maximum Ellipticity

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.753	0.734	0.745	0.789
Std Dev	0.239	0.227	0.233	0.215
Time of Day	***AM***		***PM***	
Average	0.732		0.779	
Std Dev	0.248		0.205	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.647	0.810	0.747	0.817
Std Dev	0.288	0.196	0.165	0.206
Task	***Primary***		***Primary with Secondary***	
Average	0.764		0.746	
Std Dev	0.219		0.238	

Subject (Group) was a very significant factor, $F(12,759) = 72.32$, $p < 0.001$, for Maximum Ellipticity. In addition, the Time of Day/Workload interaction showed a task dependence (Fig. 7.12). The four highest values occurred on the segments recorded during tasks performed during the approach phase regardless of Time of Day or Workload. The Nominal/AM treatments were not different from each other but were different from the Monitor/PM and High Load/PM treatments. The four lowest treatments occurred on downwind segments. Data from these four treatments were not significantly different from each other, but were significantly different from data of the four approach treatments. Maximum Ellipticity was responsive to workload changes during approach, but unresponsive to changes on downwind. Therefore, Maximum Ellipticity can be used to differentiate between task types used in this study.

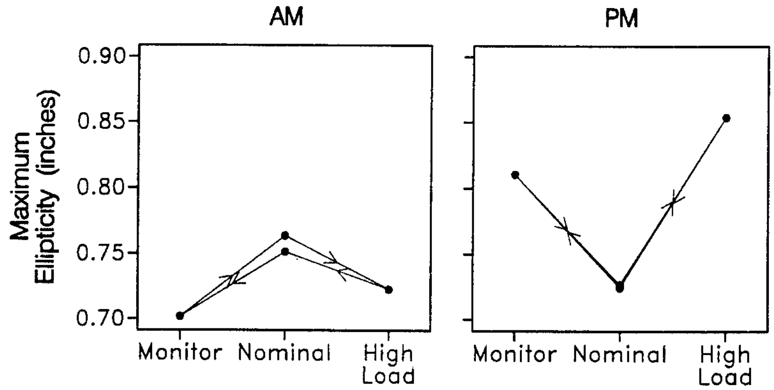

Figure 7.12. Maximum Ellipticity interaction of Time of Day/Workload

7.5.8. Maximum Ellipticity Change

The Maximum Ellipticity was measured from the center of the fixation to the most distant point included in the fixation. The change in this distance was compared to the average Maximum Ellipticity distance recorded for the last segment. Segments with a secondary task were compared to the previous segment with a secondary task. Segments without a secondary task were compared to the previous segment without a secondary task. The values for this parameter were normally distributed and ranged from –0.70 to 0.78 with an average of -0.0048. All conditions displayed uniform variance.

Only one primary factor, Workload Level, produced significant results, $F(3,36) = 8.57$, $p < 0.001$. In addition, five of the seven interactions with Workload Level also produced significant results. All five of the significant interactions included Workload Level as one of the interaction factors. Significant factors and interactions accounted for 37.9% of the ANOVA variance.

The Subject (Group) factor was not significant. This was a consistent result for "Change" variables. However, the Time of Day/Workload interaction accounted for 19% of ANOVA variance. Again, this interaction was task dependent.

Table 7.13. Factor Level Data –Maximum Ellipticity Change

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.014296	-0.070958	-0.009810	0.042160
Std Dev	0.220	0.223	0.231	0.214
Time of Day	***AM***		***PM***	
Average	-0.009556		-0.002653	
Std Dev	0.213		0.239	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	-0.010322	-0.004416	-0.004645	-0.005027
Std Dev	0.235	0.227	0.219	0.224
Task	***Primary***		***Primary with Secondary***	
Average	-0.009691		-0.002532	
Std Dev	0.206		0.245	

The morning and afternoon data sets were aligned by task type. The lower values were approach segments and the positive changes occurred on the downwind segments. The Time of Day/Workload interaction, $F(3,36) = 19.56$, $p < 0.001$, was attributable to the afternoon results where differences between downwind and approach segments was greatest.

7.6. Perception Strategy Variables

Perception Strategy is one category of parameters falling beneath the attention related parameters for arousal. The amount of attention available could theoretically create a ceiling for these parameters. If insufficient attention was available the Perception

Parameters would be affected by the ceiling created. For example, task overload occurs because of insufficient attention resources to meet a demand. Therefore, Perception Parameters should parallel Arousal Parameters for task overload since they are also attention limited.

However, if attention is not a limiting factor perception strategy can act independently to affect performance. For example, if a subject has the attention assets available to perform a task but employs a suboptimal scanning strategy, performance will suffer. This could be the case for a novice or an experienced subject who lacks the desire to perform well. Alternatively, a highly motivated subject may elect to be hypervigilant when it is not necessary to perform well.

Two of the eight Perception Strategy Parameters were not significant for the primary factor of Workload Level. The fraction of Transition Matrix Symmetric and Short Fixations had no significant factors, however Short Fixations had significant interactions which merit presentation of the parameter.

Two additional Perception Parameters, Angle Fixation Gate and Velocity Fixation Gate, were sensitive to the approach and downwind segments of the simulation. Again, it was not the intent of this study to go down to the resolution of subtasks so these parameters will not be discussed at length within the context of this study. However, these parameters merit consideration if an investigator desires to use eye movements to aid in task decomposition or adaptive display design.

Another Perception Parameter, Number of Cycles, was not developed because of numerous treatments with no valid data. It is probable the highly flexible nature of the instrument cross check precludes use of one anchor to analyze scan cycles (Zhang, 1994).

However, this parameter displayed significant correlation to both workload and performance so it is desirable to create a system by which the different viewing cycles may be analyzed. To provide valid data for all treatments it may be possible to devise a floating anchor for cycles based on a situated cognition model.

Finally, three Perception Parameters are good candidates for modeling workload and performance: Dual Fixation Gate, Transition Matrix Repeat, and Transition Matrix Useful.

7.6.1. Fraction of Velocity Fixation Gate (only)

Fixations used to calculate the Fraction of Velocity Fixation Gate (only) parameter resulted from two sources. The first fixation source was sequential fixations for two, nearby display symbols. Symbols within 1.5 inches of each other (less than one degree visual angle) were sufficiently close that only the Velocity Fixation Gate was exceeded transitioning to a new fixation. The second fixation source came from situations where the fixation duration was indicative of staring at one spot for multiple fixations. Fixations greater than one second in duration were terminated at the one second point. The Fraction of Velocity Fixation Gate was calculated as the number of fixations trapped only by the Velocity Fixation Gate divided by the total number of eye fixations (Angle Fixation Gate Only plus Velocity Fixation Gate Only plus Dual Fixation Gates).

The statistically significant factors for the Fraction of Velocity Fixation Gate were Subject(Group), $F(12,759) = 69.51$, $p < 0.001$, Time of Day, $F(1,12) = 5.12$, $p < 0.05$, and Workload Level, $F(3.36) = 3.49$, $p < 0.05$. The significant interactions were Task/Time of Day, $F(1,12) = 12.73$, $p < 0.01$, Task/Workload Level, $F(3,36) = 15.11$, $p < 0.001$,

Time of Day/Workload Level, $F(3,36) = 17.82$, $p < 0.001$, and Task/Time of Day/Workload Level, $F(3,36) = 5.45$, $p < 0.01$. The significant factors and interactions accounted for 13.4% of the ANOVA variance.

The large number of significant factors and interactions were indicative of the extreme variability of the parameter. Large numbers of Velocity Gate (only) fixations occurred in pockets. Since eye blinks were one source of these fixations, it appears short periods with a high frequency of blinks could have been responsible for the variability.

Table 7.14. Factor Level Data – Fraction of Velocity Fixation Gate

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	0.439	0.477	0.463	0.427
Std Dev	0.257	0.240	0.252	0.226
Time of Day	*AM*		*PM*	
Average	0.480		0.422	
Std Dev	0.245		0.241	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	0.552	0.400	0.459	0.393
Std Dev	0.301	0.202	0.193	0.234
Task	*Primary*		*Primary with Secondary*	
Average	0.451		0.451	
Std Dev	0.242		0.248	

The Fraction of Velocity Fixation Gate was exponentially distributed about a median of 0.268. This parameter ranged from 0.0 to 1.0 with a mean of 0.332. Heteroscedasticity was a problem for all the factors except Secondary Task. Parameter distribution and variance made it an unattractive candidate for modeling either workload or performance. The Time of Day/Workload interaction also indicated the parameter was

sensitive to segment type (approach or downwind). The Velocity Gate parameter could be useful differentiating task type.

7.6.2. Fraction of Angle Fixation Gate

Fixations categorized as Angle Fixation Gate have two sources. The first source was tracking of slow moving objects which would not trip the saccadic velocity gate. The second source was fixations occurring immediately after an eye blink. The invalid tracking status caused by a blink disabled the velocity gate. In these two cases, only the Angle Gate would identify the fixations. Since eye blinks are an opportunistic mechanism (Gray, 1977; Skelly, 1993), they would occur most often when the visual scan pattern is not heavily taxed, and in the case of fatigue (Stern, 1987). Likewise, tracking of objects would occur only when the subject perceived time was available to complete the tracking task before moving on to another task. The Fraction of Angle Fixation Gate was calculated by dividing the number of fixations trapped only by the Angle Fixation Gate by the total number of eye fixations (Angle Gate Only plus Velocity Gate Only plus Dual Fixation Gates). It was expected the Fraction of Angle Fixation Gate would decrease with increased workload.

Task, $F(1,12) = 9.42$, $p < 0.01$, Subject(Group), $F(12,759) = 69.49$, $p < 0.001$, Time of Day, $F(1,12) = 5.61$, $p < 0.05$, and Workload Level, $F(3,36) = 5.55$, $p < 0.01$, were all significant Main Effect factors for the parameter Angle Fixation Gate. The Interactions Task/Workload Level, $F(3,36) = 10.65$, $p < 0.001$, Subject(Group)/Workload Level, $F(9,36) = 2.23$, $p < 0.05$, Time of Day/Workload Level, $F(3,36) = 14.03$, $p < 0.001$, and Task/Time of Day /Workload Level, $F(3,36) = 4.50$, $p < 0.01$, were all

significant. Significant factors and interactions accounted for 14.4% of the ANOVA variance. Data was distributed normally but the variance was not homogenous for the Time of Day factor.

Table 7.15. Factor Level Data – Fraction of Angle Fixation Gate

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.346	0.302	0.304	0.294
Std Dev	0.240	0.228	0.242	0.226
Time of Day	***AM***		***PM***	
Average	0.285		0.338	
Std Dev	0.222		0.244	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.282	0.347	0.281	0.336
Std Dev	0.271	0.221	0.185	0.246
Task	***Primary***		***Primary with Secondary***	
Average	0.270		0.353	
Std Dev	0.228		0.235	

The Angle Gate (only) increased with the addition of a secondary task, $p < 0.001$. This factor was the most dominant, accounting for 4.1% of the ANOVA variance. For Workload, the Monitor treatment was 4% higher, $p < 0.001$, for Fraction of Angle Gate than the three other treatments, Nominal After High Load, Nominal After Monitor and High Load. The other three treatments were not significantly different from each other.

During the Monitor segments, subjects appeared to use an excessive amount of time tracking the aircraft figure on the horizontal situational display. This fact indicates the Angle Fixation Gate (only) variable differentiated between working scan patterns and observational scan patterns quite well. Thus, Fraction of Angle Fixation is higher when the task is easy.

Two of the interactions seen with most other parameters were also present here. The Time of Day/Workload interaction was aligned by segment type (downwind or approach). The Task/Time of Day/Workload interaction displayed the usual shift in viewing strategy for the Primary Task/PM/Nominal After High Load treatment. These interactions were typical of the perception strategy parameters.

7.6.3. Fraction of Dual Gate Fixations

Two methods were used to identify fixations within the raw eye tracking data. Both methods were used simultaneously as described in Chapter 4. In many cases, the same data point triggered both the saccadic velocity gate and the angle cut-off gate simultaneously. This parameter, Fraction of Dual Gate Fixations, tracked the fraction of fixations trapped by both methods as a function of the total number of fixations in the 36 second segment.

A Dual Gate fixation was indicative of deliberate movement among visual targets, devoid of eye blinks and pursuit (tracking) movement. Deliberateness of the scan pattern was not necessarily an indication of either good or bad viewing strategy. The fraction of Dual Gate fixations could have been increased if the subject was a novice performing free scanning or an expert with much flexibility built into his/her scan pattern.

The Fraction of Dual Fixation Gate was calculated by dividing the number of Dual Fixation Gates by the total number of eye fixations (Angle Gate Only plus Velocity Gate Only plus Dual Fixation Gates). Two factors, Time of Day and Workload Level, and their interaction displayed uniform variance.

Table 7.16. Factor Level Data – Fraction of Dual Gate Fixation

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.224	0.232	0.243	0.289
Std Dev	0.107	0.109	0.119	0.132
Time of Day	***AM***		***PM***	
Average	0.245		0.248	
Std Dev	0.118		0.121	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.184	0.260	0.262	0.282
Std Dev	0.094	0.116	0.109	0.132
Task	***Primary***		***Primary with Secondary***	
Average	0.290		0.204	
Std Dev	0.119		0.104	

The percentage of fixation identified by both methods ranged from 0% to 78% with an average of 24.7%. Standard deviation for the population was 12%. Subject(Group), $F(12,759) = 21.22$, $p < 0.001$, and Workload Level, $F(3,36) = 19.92$, $p < 0.001$, were the significant factors for this parameter. The significant interactions were Task/Group, $F(3,12) = 3.76$, $p < 0.05$, and Time of Day/Workload Level, $F(3,36) = 8.21$, $p < 0.001$. Significant factors and interactions accounted for 32.3% of the ANOVA variation.

Dual Fixation Gates fixations were more prevalent when subjects experienced a higher workload as with the Primary Task treatments. (Performance results had previously demonstrated the Primary Task treatments were the higher workload segments, $p < 0.001$.) The Fraction of Dual Fixation Gates increased for the Primary Task treatment, when a focused instrument cross check became necessary to return the simulator to required instrument conditions.

Figure 7.12 illustrates the progression of Percentage of Dual Gate Fixations as Workload Level increases. The High Load treatment was approximately 5% greater than all other treatments, $p < 0.001$. Other treatments were within 1% of each other, and were not significantly different. Higher workload resulted in a more deliberate instrument cross check.

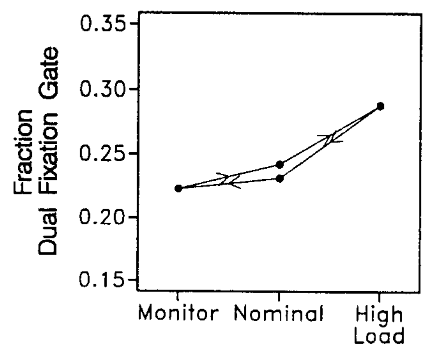

Figure 7.13. Fraction of Dual Gate Fixations versus Workload

A Task/Group interaction occurring with this parameter was a result of the Novice group. All other groups recorded a significantly lower fraction of Dual Gate fixations, $p < 0.001$, with Primary and Secondary Task. Dual Fixation Gate numbers increased with workload, acting as an indicator of the deliberateness of the scan strategy. However, it is not a valid metric if no scanning strategy exists as might be the case with some novices.

The interaction of Time of Day and Workload Level was significant because of one data point. The PM treatment for the Nominal After Monitor Workload Level was significantly lower than its AM counterpart. Figure 7.14. illustrates the anomaly. Previously, the Nominal After High Load factor level produced this interaction.

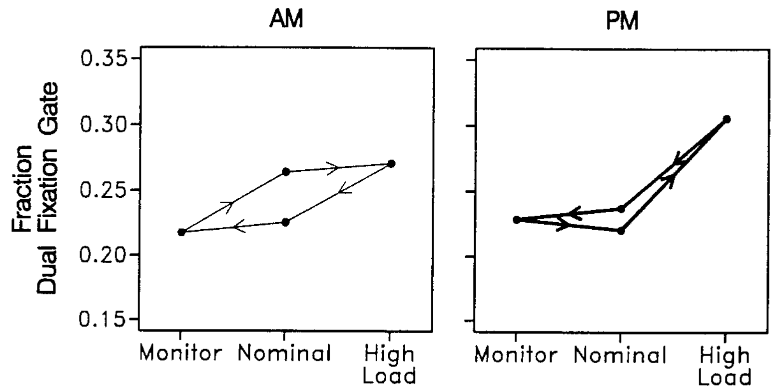

Figure 7.14. Fraction Dual Gate interaction of Time of Day/Workload

7.6.4. Percent Transition Matrix Symmetric

Transition matrix symmetry is related to the nature of the viewing task. Free viewing tasks have demonstrated symmetric transition matrices (Biederman et al, 1981), whereas structured tasks naturally result in specific transition patterns causing matrix asymmetry. Fixations were assigned to areas of interest (described in Chapter 4) and eye scan patterns were recorded among the areas of interest. Asymmetric fixations were those fixations in any given matrix column/row combination without a compliment in the same row/column combination. The percentage of transition matrix symmetry was calculated by dividing the symmetric fixations by the total number of fixations. The percentage of matrix symmetry ranged from 63% to 100% with an average of 84%. The percentage was normally distributed around the median of 85.7%. The standard deviation was 8.25%. Interactions accounted for 17.3% of the ANOVA variance.

Table 7.17. Factor Level Data – Percent Transition Matrix Symmetric

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	86.2	85.2	84.9	83.8
Std Dev	7.8	8.6	8.5	8.1
Time of Day	*AM*		*PM*	
Average	85.3		84.7	
Std Dev	8.1		8.5	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	88.3	83.0	84.9	83.8
Std Dev	8.5	7.9	7.0	8.6
Task	*Primary*		*Primary with Secondary*	
Average	85.3		84.7	
Std Dev	8.46		8.07	

Workload was not a significant factor. The factor Subject(Group), $F(12,759) = 17.57$, $p < 0.001$, and the interactions Time of Day/Workload Level, $F(3,36) = 22.26$, $p < 0.001$, and Task/Time of Day/Workload Level, $F(3,36) = 7.18$, $p < 0.001$ showed significance.

Symmetry results were unremarkable with one exception. Like other strategy variables, there was a strong Task/Time of Day/Workload interaction corresponding to a unique performance error increase in the afternoon. Other interactions were related to the segment type (approach versus downwind). The Primary Task/PM/Nominal After High Load treatment resulted in the highest percentage of matrix symmetry of all three factor combinations. Subjects were free scanning in lieu of using their more structured instrument cross check scanning. This behavior would indicate subjects were aroused but scanning the wrong indicators during the Primary/PM/Nominal After High Load performance/vigilance decrement.

7.6.5. Percent Transition Matrix Repeat Fixations

Fixations were assigned to numbered areas of interest and eye scan patterns were recorded among the areas of interest. Repeat fixations within specified areas of interest were tracked as a metric of viewing pattern efficiency. When sequential fixations fell within the same area of interest they were considered to be repeat fixations. The percentage of repeat fixations was calculated by dividing the number of repeat fixations by the total number of fixations. The percentage of repeat fixations ranged from 3.6% to 98.3% with an average of 48.9%. Repeat fixation data was normally distributed around a median of 47.2%. The standard deviation was 14.5%. Significant factors and interactions accounted for 8% of the ANOVA variance. The five factor ANOVA may be found in Appendix E, Table E24.

Workload Level, $F(3,36) = 10.71$, $p < 0.001$, and the interactions Task/Workload, $F(3,36) = 4.56$, $p < 0.01$, Time of Day/Workload, $F(3,36) = 5.18$, $p < 0.01$, and Task/Time of Day/Workload, $F(3,36) = 3.21$, $p < 0.05$, were significant. Repeat fixations decrease with the increase in Workload Level. Only the High Load treatment was significantly different from all other treatments, $p < 0.01$. The three factor interaction again displayed a change in strategy occurring for the Primary Task/PM/Nominal After High Load treatment.

Table 7.18. Factor Level Data – Percent Matrix Repeat Fixations

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	51.2	50.8	49.0	44.9
Std Dev	14.6	14.2	14.0	14.7
Time of Day	*AM*		*PM*	
Average	49.5		48.4	
Std Dev	15.3		13.8	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	56.8	45.8	48.3	45.0
Std Dev	17.3	12.5	10.9	13.6
Task	*Primary*		*Primary with Secondary*	
Average	49.7		48.2	
Std Dev	14.9		14.2	

7.6.6. Percent Matrix Useful

Percent Matrix Useful was designed to represent the percentage of fixations that resulted in acquisition of information useful to completion of the required tasks. This parameter included both the primary task and the secondary tasks. A fixation was considered useful if it fell within a useful area of interest. Repeat fixations were considered useful if they were less than the fifth consecutive fixation in that area. The percentage of Useful Fixations ranged from 4.0% to 96.3% with and average value of 73.6%. Data was not uniformly distributed. Group and Time of Day factors exhibited heteroscedasticity, even though these factors were significant.

Table 7.19. Factor Level Data – Percent Matrix Useful

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	67.5	73.4	74.5	78.9
Std Dev	17.9	16.4	17.0	16.8
Time of Day		*AM*		*PM*
Average		73.3		73.9
Std Dev		19.6		15.0
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	64.2	76.9	75.1	78.2
Std Dev	22.4	13.9	13.6	14.8
Task		*Primary*		*Primary with Secondary*
Average		73.7		73.5
Std Dev		17.8		17.2

The significant factors and interactions for Percent Matrix Useful were Subject(Group), $F(12,759) = 23.12$, $p < 0.001$ and Workload Level, $F(3,36) = 20.43$, $p < 0.001$. Time of Day/Workload Level, $F(3,36) = 7.02$, $p < 0.001$, and Task/Group/Workload, $F(3,12) = 7.43$, $p < 0.01$, were the significant interactions. This rare Group interaction was due to very low data for the Primary Task/Novice/AM treatment. The novices did not know where to look during the high load, morning segments.

Figure 7.15 illustrates the bimodal distribution for Percent Useful Fixations. At approximately 40% Useful Fixations all performance was nominal. Above and below that figure variance increased as the range of performance error increased into an area where ATC would be engaged to prevent accidents.

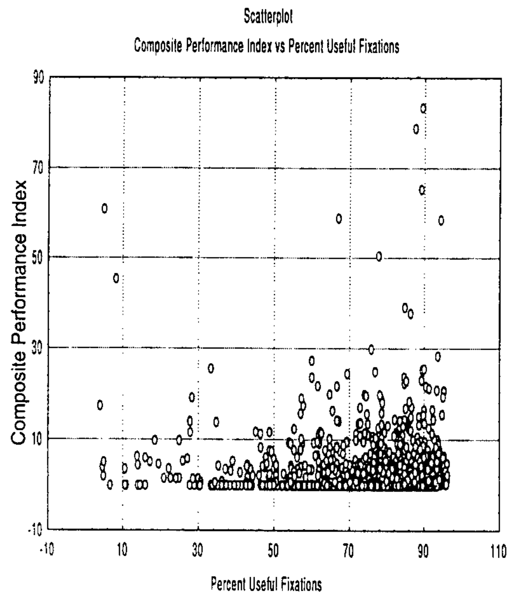

Figure 7.15. Performance versus Percent Useful Fixations

Workload Level was a significant factor as well. Percent Useful Fixation consistently increased across the levels of Workload. The two Medium levels were not significantly different from each other, but they were significantly different from both the Monitor and High Load treatments, $p < 0.01$. This progression and the significant differences among factor levels match the results of Performance Error Rating.

The Time of Day/Workload Level interaction revealed that subjects increased the usefulness of their scan patterns by 6% at the PM/High Load Level. This increase was not significant, yet the increase in usefulness corresponds to better performance.

7.6.7. Short Fixations

Short fixations were those fixations less than 0.2 seconds in duration. The total number of short fixations per 36 second data segment was entered as the parameter, Short Fixations. The number of short fixations per 36 second data segment ranged from 3 to 74

with an average of 24.7 and a standard deviation of 9.58. The number of Short Fixations per segment was normally distributed around the median of 24. Significant Interactions accounted for 15.5% of ANOVA variance.

Table 7.20. Factor Level Data – Short Fixations

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	26.1	24.5	24.0	24.2
Std Dev	9.8	9.5	9.6	9.3
Time of Day	***AM***		***PM***	
Average	24.2		25.2	
Std Dev	9.3		9.8	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	24.0	24.1	26.3	24.3
Std Dev	9.6	8.4	9.9	10.2
Task	***Primary***		***Primary with Secondary***	
Average	23.6		25.8	
Std Dev	9.4		9.6	

The only significant factor for Short Fixation was Subject(Group), $F(12,759) = 24.90$, $p < 0.001$. Neither Workload nor its commonly occurring interaction with Time of Day were significant. However, the Task/Time of Day/Workload, $F(3,36) = 6.77$, $p < 0.01$, interaction accounted for 2.8% of ANOVA variance. Figure 7.16 shows the unique results for Primary Task/PM/Nominal After High Load. This low value among all treatments was the same for numerous other three factor interactions where the increase in performance error was seen. The number of Short Fixations for this particular treatment was 17% lower than its nearest Nominal After High Load counterpart. Composite Performance Error for this same treatment was 80% higher than the nearest Nominal After High Load counterpart.

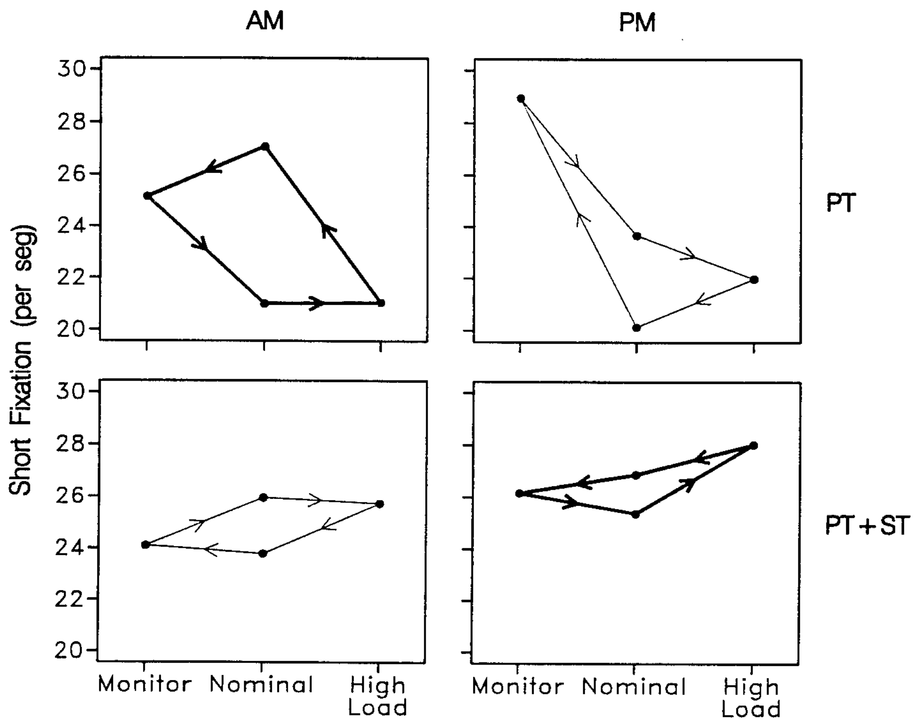

Figure 7.16 Short Fixation interaction of Task/Time of Day/Workload

7.6.8. Number of Cycles

Viewing Cycles were anchored on the altimeter, the most visited area of interest. A cycle was counted any time the subject's eye scan returned to the altimeter area of interest.

Subject(Group), $F(12,759) = 28.18$, $p < 0.001$, and Workload Level, $F(3,36) = 8.23$, $p < 0.001$, were the significant factors. Task/Workload Level, $F(3,36) = 4.11$, $p < 0.05$, Time of Day/Workload Level ($p<0.001$), and Task/Time of Day/Workload Level, $F(3,36) = 14.07$, $p < 0.001$, were the significant interactions.

The Number of Cycles results were encouraging, since the significant factors and interactions mirrored performance error. However, numerous treatments had a zero average (no cycles) spanning four replicates so statistical analysis comparable to that for the previous parameters was not possible. In addition, the Group and Workload Level data for the five "Cycle" related parameters was heteroscedastic.

7.7. Cognitive Processing Parameters

As cognitive load increases, the frequency of fixations decrease and the duration of fixations increase (Rayner and Morris, 1990; Williams and Harris, 1985; Just and Carpenter, 1976). The cause of cognitive loading may be a primary task, secondary tasks, or distraction but the result is the same. This section will present results related to length of fixation as related to cognitive processing and an EEG index conceived to measure engagement as a metric for cognitive processing.

7.7.1. Fixation Time

Fixation time was started at the first eye tracking point falling within both the saccadic velocity gate and the angle tracking gate. Fixations were terminated at the last point before one or both of the gates were exceeded. Average fixation time was computed for each data segment. Data on individual fixations was also analyzed to determine the number of Short Fixations and Long Fixations. Long and short fixations where tracked to determine whether there were general or specific causes in fixation time.

Data was normally distributed with a median value of 0.319 sec and a standard deviation of 0.049 sec. The average fixation time for the following factors ranged from

0.305 to 0.324 seconds. This range compares well with the generally accepted average fixation duration of approximately 300 milliseconds (Just and Carpenter, 1976; Harris and Glover, 1985). A total of 73,982 fixations were recorded over 1016 data segments. An average of 2.02 fixations per second was recorded.

Subject(Group) was a significant factor, $F(12,759) = 44.15$, $p < 0.001$. Task, $F(1,12) = 5.97$, $p < 0.05$, and Workload Level, $F(3,36) = 11.39$, $p < 0.001$, were the other significant factors. The significant interactions were Task/Workload Level, $F(3,36) = 7.38$, $p < 0.001$, and Time of Day/Workload Level, $F(3,36) = 6.69$, $p < 0.001$. Significant factors and Interactions accounted for 14.1% of the ANOVA variance.

Table 7.21. Factor Level Data – Fixation Time

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.304	0.317	0.318	0.324
Std Dev	0.041	0.045	0.047	0.044
Time of Day	***AM***		***PM***	
Average	0.319		0.313	
Std Dev	0.045		0.044	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.311	0.310	0.324	0.318
Std Dev	0.044	0.048	0.033	0.051
Task	***Primary***		***Primary with Secondary***	
Average	0.324		0.308	
Std Dev	0.045		0.043	

As expected, less demanding tasks decreased Fixation Time (Just and Carpenter, 1976). Fixation time for the Primary Task treatments averaged 15 milliseconds more, $p < 0.05$, than the Primary with Secondary Task treatment. Composite performance variables had a similar Task effect. The difference among Workload factors spanned 20

milliseconds, a 7% increase. Figures depicting these changes will be presented with the Long Fixation parameter since these two parameters were very similar, but the Long Fixation Parameter had less variance.

Fixation Time increased with increasing Workload Level. This parallels the increase in Performance Error Rating. The Monitor treatment of Fixation Time was approximately 13 milliseconds different from treatments for Nominal After High Load, $p < 0.05$. Differences between the other Workload factors was even greater. Like the performance rating, the two Nominal treatments were not significantly different.

Fixation Time interactions were similar to Long Fixation in significance, sense, and form. However, Long Fixation had a greater correlation to workload and performance. The interactions significant for Fixation Time will be presented in greater detail with the Long Fixation parameter.

7.7.2. Fixation Time Change

Fixation Time Change was computed as the difference between the current segment Fixation Time and the Fixation Time for the previous segment with like Task treatment. Values for Fixation Time Change were normally distributed. The values ranged from –186ms to 203ms and standard deviation was 1.5 ms.

The factors of Task, $F(1,12) = 8.93$, $p < 0.05$, Time of Day, $F(1,12) = 8.58$, $p < 0.05$, and Workload Level, $F(3,36) = 7.49$, $p < 0.001$, were significant and possessed uniform variance. The Task/Workload Level, $F(3,36) = 4.77$, $p < 0.01$, and Time of Day/Workload Level, $F(3,36) = 5.24$, $p < 0.01$, interactions were significant as well. Significant factors and interactions accounted for 17.9% of ANOVA variance.

Fixation Time Change increased, $p < 0.05$, when there was no secondary task. Performance Results show changes associated with this factor were related to the subjects attempts to minimize error before completing the secondary task. The increase in Fixation Time Change was related to problem solving on the aviation task.

In Table 7.22, Fixation Time Change was only 7 milliseconds different between Time of Day treatments but the difference was significant, $p < 0.05$. In the morning, Fixation Times decreased slightly through the simulation period, indicating an improvement in instrument cross check efficiency through the simulation. No other factors or interaction displayed recognizable trends.

Table 7.22. Factor Level Data –Fixation Time Change

Workload	*Monitor*	*Nom(After High)*	*Nom(After Mon)*	*High Load*
Average	0.014341	-0.021802	-0.014007	-0.014687
Std Dev	0.110	0.115	0.119	0.110
Time of Day	*AM*		*PM*	
Average	-0.005064		0.001655	
Std Dev	0.107		0.122	
Group	*Novice*	*NASA Tech*	*Air Force Pilot*	*Commercial Pilot*
Average	-0.003725	-0.002837	-0.000050	-0.000266
Std Dev	0.125	0.123	0.101	0.107
Task	*Primary*		*Primary with Secondary*	
Average	-0.003524		0.000098	
Std Dev	0.104		0.124	

7.7.3. Long Fixations

Long Fixations were an indicator of decreased visual processing and increased cognitive processing for problem solving. In some cases, this was due to increased cognitive activity related to the previous or current fixation (Just and Carpenter, 1986;

Harris et al., 1992). Long fixations were those fixations more than one standard deviation longer in duration than the average Fixation Time. A running tally of Long Fixations was kept for each data segment, and the final tally was recorded at the end of the segment as a segment variable. The average number of Long Fixations per 36 second segment was 21 with a standard deviation of 8.5. The number of Long Fixations per segment ranged from 0 to 50. Parameter values were normally distributed around a median of 22.

Task, $F(1,12) = 7.31, p < 0.05$, Workload Level, $F(3.36) = 16.3, p < 0.001$, and three interactions had statistically significant differences among their treatment levels. Mean values and standard deviations for each factor level may be found in Table 7.23.

Table 7.23 illustrates the different number of Long Fixations with and without a Secondary Task. The spread between the two treatments was approximately three fixations or a 13% decrease with addition of a secondary task. Performance Error Rating decreased 11% with the addition of the secondary task.

Table 7.23. Factor Level Data – Long Fixation

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	18.5	21.4	21.8	23.0
Std Dev	7.6	8.3	9.0	8.6
Time of Day	***AM***		***PM***	
Average	21.9		20.4	
Std Dev	8.3		8.7	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	20.0	19.8	23.7	21.3
Std Dev	8.6	8.2	6.9	9.6
Task	***Primary***		***Primary with Secondary***	
Average	22.7		19.6	
Std Dev	8.9		7.9	

The five factor ANOVA for Long Fixations may be found below in Table 7.24. Significant factors and interactions accounted for 15.6% of ANOVA variance.

Table 7.24 Five Factor ANOVA for Long Fixations

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	634.15	634.15	T x S(G)	7.31*	0.043
Group	[G]	3	640.82	213.61	S(G)	0.50	NS
#Subject(Group)	[S(G)]	12	20431.61	1702.63		54.33***	
Time Day	[D]	1	134.54	134.54	D x S(G)	2.55	NS
Workload Level	[W]	3	722.08	240.69	W x S(G)	16.30***	0.053
Interactions							
T x G		3	102.83	34.28	T x S(G)	0.40	NS
T x D		1	5.13	5.13	T x D x S(G)	0.46	NS
T x W		3	315.44	105.15	T x W x S(G)	7.81***	0.021
G x D		3	49.39	16.46	D x S(G)	0.31	NS
G x W		9	154.03	17.11	W x S(G)	1.16	NS
D x W		3	419.35	139.78	D x W x S(G)	9.18***	0.029
T x G x D		3	53.99	18.00	T x D x S(G)	1.60	NS
T x G x W		9	235.67	26.19	T x W x S(G)	1.94	NS
T x D x W		3	148.29	49.43	T x D x W x S(G)	3.70*	0.008
G x D x W		9	90.95	10.11	D x W x S(G)	0.66	NS
T x G x D x W		9	120.53	13.39	T x D x W x S(G)	1.00	NS
Error Terms							
S(G)		12	5107.90	425.66			
T x S(G)		12	1040.92	86.74			
D x S(G)		12	632.64	52.72			
W x S(G)		36	531.64	14.77			
T x D x S(G)		12	134.60	11.22			
T x W x S(G)		36	484.69	13.46			
D x W x S(G)		36	548.30	15.23			
T x D x W x S(G)		36	481.25	13.37			
Total		255	12789.13				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Figure 7.17 illustrates the systematic increase in the number of Long Fixations as Workload Level increases. This increase parallels that for performance error and like performance error, the Nominal treatments were not different. However, unlike

performance error results, the High Load treatment was not significantly different from the Nominal treatments.

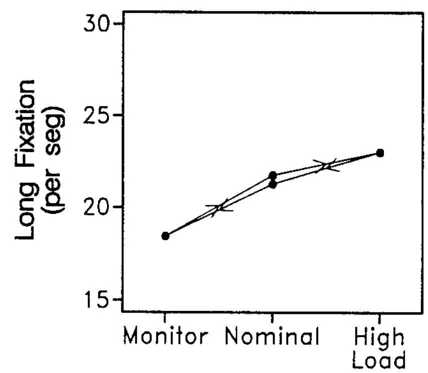

Figure 7.17. Long Fixations per Segment versus Workload Level

Performance Error Rating started at the baseline of zero for the Monitor treatment, increased to 0.70 for the Medium treatments, and increased an additional 0.65 (93%) for the High Load treatment. Starting at a baseline of 18.5 Long Fixations for the Monitor treatment, there was an increase of three fixations to the Medium treatments and an additional increase of 1.5 Long Fixations (50%) to the High Load treatment.

An illustration of the interaction of Time of Day/Workload Level for Long Fixation is in Figure 7.18. There are two data points accounting for the interaction. First, the Monitor treatment in the morning was comparable to the Medium treatments. Second, the Nominal After High Load treatment in the afternoon was most comparable to the High Load treatments. The clockwise progression of long fixation data reversed itself in the afternoon due to the change in position for the Nominal values.

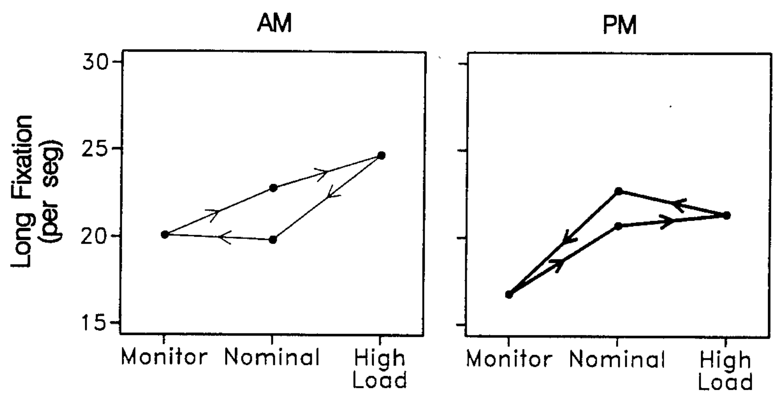

Figure 7.18. Long Fixation interaction of Time of Day/Workload

The other significant two factor interaction for Long Fixation was Task/Workload Level. Figure 7.19 illustrates the very regular progression for the afternoon task. The Monitor treatment in the morning again proved different, $p < 0.001$, with respect to the other three morning treatments. The interaction occurs because the Primary with Secondary Task treatments showed only small increases with increased workload.

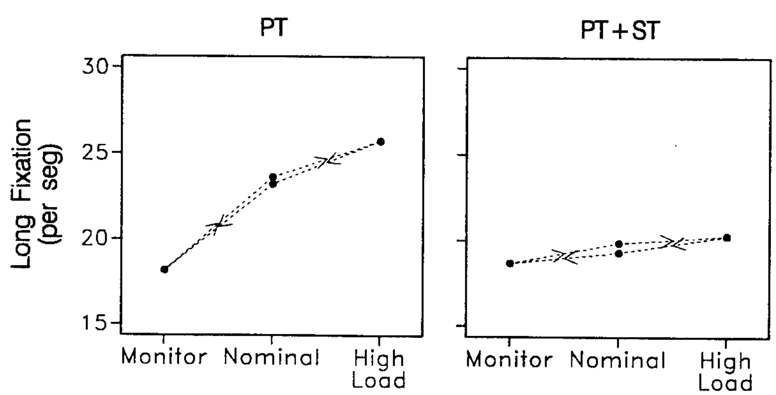

Figure 7.19. Long Fixation interaction of Task/Workload Level

The three factor interaction of Workload Level, Time of Day, and Task (Figure 7.20) illustrates the expected interaction at the Nominal After High Load treatment. The morning baseline conditions resulted in approximately 20 long fixations per data segment. Most afternoon treatment variations resulted in a level or decreasing number of long fixations compared to morning. The lone exception was the afternoon treatment with only a primary task. The number of long fixations increased by 30%. This same treatment consistently created the 3-Way interaction for eye movement parameters in which the three factor interaction was found.

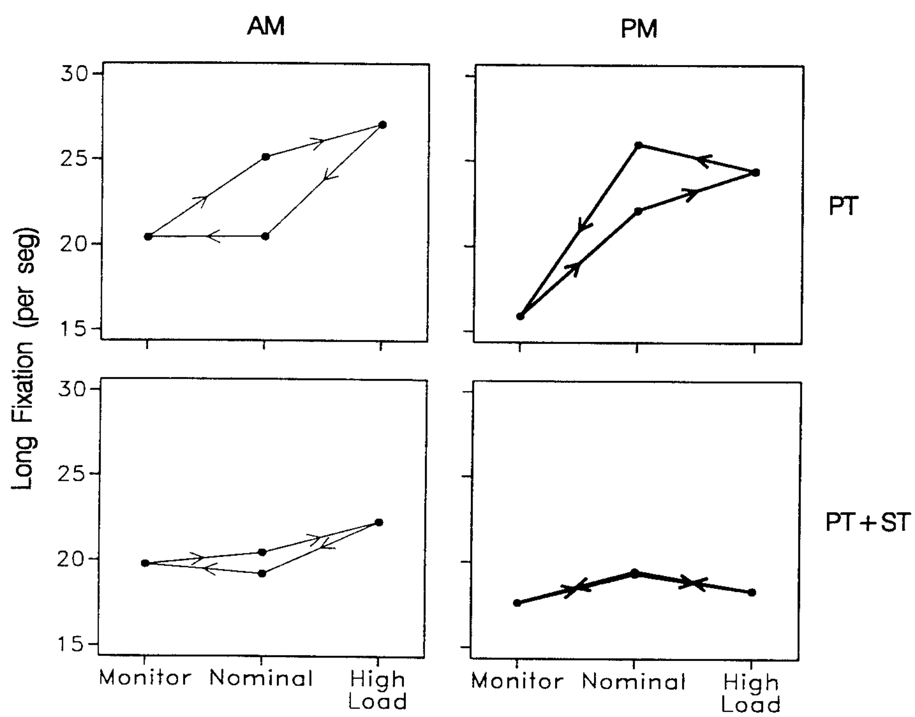

Figure 7.20. Long Fixation interaction of Task/Time of Day/Workload Level

7.7.4. Index of Engagement

Index of Engagement was a ratio of the power spectral densities for three EEG frequency elements, $\beta/(\alpha+\Theta)$. Index of Engagement was recorded at 2 Hz and averaged over the 36 second segment period. An increase in the index can result from either an increase in high frequency EEG (β), or a decrease in lower frequency EEG (α or θ). Values for Index of Engagement ranged from 0.092 to 1.895 with a mean of 0.771. Data was normally distributed with a standard deviation of 0.290.

Index of Engagement decreased 16% with the addition of a secondary task. Given the unexpectedly good performance results associated with the Secondary Task treatment it was not surprising that the index was lower with a secondary task. Treatments with a secondary task were easier because subjects had minimized performance error before attempting the secondary task.

Table 7.25. Factor Level Data – Index of Engagement

Workload	***Monitor***	***Nom(After High)***	***Nom(After Mon)***	***High Load***
Average	0.792	0.770	0.777	0.713
Std Dev	0.296	0.300	0.288	0.270
Time of Day	***AM***		***PM***	
Average	0.758		0.768	
Std Dev	0.293		0.287	
Group	***Novice***	***NASA Tech***	***Air Force Pilot***	***Commercial Pilot***
Average	0.827	0.782	0.600	0.842
Std Dev	0.350	0.250	0.189	0.281
Task	***Primary***		***Primary with Secondary***	
Average	0.831		0.692	
Std Dev	0.280		0.284	

However, the High Load treatment was significantly lower than all other treatments, $p < 0.001$. A decrease of Index of Engagement was not expected with an increase of workload. Neither did interactions follow any expected or discernable pattern with respect to the significant interactions of Group/Workload, Time of Day/Workload, or Group/Time of Day/Workload.

The irregular data patterns leave some cause for question about the usefulness of data as it was processed here. The total difference across workload levels was 0.079 with a minimum standard deviation among individual workload factors of 0.270. Workload accounted for 1.1% of total ANOVA variance while Task accounted for 7.1%.

7.8. Summary of Psychophysiological Parameters

Seventeen of the twenty-four parameters were related to Workload. However, not all of the 17 parameters are good candidates for modeling Performance Error which demonstrated a high correlation to Workload and Workload interactions with Time of Day and Task. Several elements need to be considered to determine the appropriateness of these parameters for modeling performance. Those elements include: 1) significance of the Workload factor, 2) significance of other primary factors, 3) significance of interactions similar/dissimilar to performance results, 4) ANOVA variance accounted for, 5) heteroscedasticity, and 6) factor independence. The first two elements, significance of Workload and other primary factors, will be summarized from Table 7.1. Heteroscedasticity was also summarized in Table 7.1. Similarity of interactions and ANOVA variance are summarized in Table 7.26. Factor independence and correlation to performance will be considered below.

Table 7.26. Summary of Significant Correlations to Workload

Psychophysiological Parameter	Correlates to Workload / (Performance)	Time Day/DL Interaction Correlates	2nd Task/TOD/DL Interaction Correlates	ω^2
Arousal (Attention) Parameters				
Pupil Diameter	P<0.001	NS	NS	0.044
Pupil Diameter Change	P<0.001	NS	NS	0.252
Peripheral Temp	P<0.001	NS	NS	0.293
Change Peripheral Tmp	P<0.001	NS	NS	0.659
Sensory Parameters				
Saccade Time	P<0.001	Yes	NS	0.174
Saccade Time Change	P<0.01	NS	NS	0.097
Saccade Distance	NS	NS	NS	0.000
Saccade Dist Change	NS	NS	NS	0.000
Fixation Size	NS	Segment	Yes	0.063
Fixation Size Change	p<0.01	Segment	NS	0.111
Maximum Ellipticity	p<0.05	Segment	Yes	0.126
Max. Ellipticity Chng	p<0.001	Segment	NS	0.379
Perception (Strategy) Parameters				
Velocity Fix Gate	p<0.01	NS	Yes	0.155
Angle Fix Gate	p<0.05	Segment	Yes	0.142
Dual Fixation Gate	p<0.001	Segment	Yes	0.134
Trans Matrix Symmetry	NS	NS	Yes	0.323
Trans Matrix Repeat	p<0.001	Segment	Yes	0.173
Trans Matrix Useful	p<0.001	NS	Yes	0.080
Short Fixations	NS	NS	Yes	0.117
Number of Cycles	p<0.001	Yes	Yes	N/A
Cognitive Parameters				
Fixation Time	p<0.001	Yes	Yes	0.141
Fixation Time Change	p<0.001 (NP)	NS	NS	0.179
Long Fixations	p<0.001	Yes	Yes	0.156
Index of Engagement	p<0.001 (NP)	NS	NS	0.096

NS – Not Significant, NP – No Parallel Data Trends,
Segment – Results Aligned with Approach or Downwind Segment

7.8.1. Selection for Performance Modeling - Parameters related to Arousal

Pupil diameter is associated with arousal (Kahnemàn, 1973). In addition, this study demonstrated the correlation of the peripheral temperature variable with workload. However, neither variable displayed interactions corresponding to the performance and psychophysiological parameters. Instead, the Parameters showed great similarity to each other.

The Pupil Diameter Change and Peripheral Temperature Change variables were also similar and they did show the desired interactions. Measuring the change in pupil diameter and peripheral temperature was an effective way to reduce the variability of the parameters. Pupil Diameter Change and Peripheral Temperature Change both had reduced heteroscedasticity compared to the data of the basic measurements of Pupil Diameter and Peripheral Temperature. The reductions in variance and significant interactions were accompanied by large increases in the amount of variance accounted for by significant factors of the ANOVA. Peripheral Temperature accounted for 65.9% of ANOVA variance, making it a strong candidate for modelling performance. Peripheral Temperature was also a more significant and less variable index compared to Pupil Diameter. This was expected given the numerous reactive mechanisms associated with change of pupil diameter (Stern, 1987; Gray, 1977).

Peripheral temperature also provided some ability to gauge the level of arousal with which the subjects entered the simulation study. The commercial pilots were very wary of simulators since they used simulators to train and take check rides for emergency procedures. They expected something to happen when there was a drop off in aviation workload because simulator instructors commonly initiated emergency procedures under

these circumstances. Therefore, as workload dropped off for the Nominal After High Load treatments commercial pilots became more wary. This difference for the commercial pilot group was previously shown in the peripheral temperature results.

One commercial pilot entered the morning simulation period convinced it would be a test of his ability to react to emergencies. Through the morning simulation his peripheral temperature was low and covered only a two degree range. Approximately, 0.5 hours into the afternoon simulation he asked, "You really aren't going to give me EPs (Emergency Procedures), are you?" Within the next twelve minutes his peripheral temperature rose 14° F. Performance error also increased moderately after he relaxed.

7.8.2. Selection for Performance Modeling - Early Perception

Psychophysiological parameters related to this stage were those describing the basic eye movement characteristics that were independent of scan pattern and cognitive processing. Sensory Function parameters describe how subjects look at something, not where they look, or how they process the information gleaned. Saccadic measures are included in this section since little or no cognitive processing takes place during saccadic movement (Biederman, 1991).

Temporal resolution of the oculometer was 33 ms. Knowing the average saccade time would be only slightly longer than the resolution of the oculometer it was considered unlikely there would be any significance to the Saccade Time or Saccade Time Change parameters. In addition, the fixation analysis code would have to be very exact in determining fixation start and end points otherwise saccades could easily be absorbed into

adjoined fixations. Fortunately, the oculometer and fixation code performed well enough to provide useful saccadic information.

The significance of numerous factors and interactions in the Saccade Time ANOVA, and the similarity to fixation time results raised an alternative hypothesis for Saccade Time. The hypothesis was that the results might be an byproduct of the eye movement analysis code mirroring fixation time. If this were the case, all results of saccade time would most likely be a compliment to those of fixation time. Both parameters were significant for factors of Secondary Task, Subject, and Workload Level.

However, the form of the significance varied between Fixation Time and Saccade Time, as did the levels of significance. Furthermore, Saccade Time did not possess as many or the same interactions as Fixation Time. The smaller number of Saccade Time interactions was expected, since Saccade Time would not reflect additional time for cognitive processing. Saccade Time was a predictable, significant parameter with significant factors and interactions accounted for 17.4% of ANOVA variance. This was slightly better than the 14.1% accounting for by Fixation Time factors and interactions.

7.8.3. Selection for Performance Modeling - Perception (Strategy)

Perception is the first of two parameter categories falling beneath the attention cloud within the Modified HIP Model. Two of the eight Perception Parameters were not significant for the primary factor of Workload Level. The fraction of Transition Matrix Symmetric and Short Fixations had no significant factors. However, Short Fixations had significant interactions.

Three other Perception Parameters were good candidates for modeling workload and performance. Dual Fixation Gate, Transition Matrix Repeat, and Transition Matrix Useful correlated to workload and performance and had no Time of Day/Workload Level interaction. The absence of the two factor interaction was important because it showed viewing strategies used by subjects were consistent between morning and afternoon except in one specific three factor interaction.

The Secondary Task/Time of Day/Workload level interaction occurring at the Primary Task/PM/Nominal After High Load treatment demonstrated a shift in viewing strategy occurring under very specific conditions. After lunch subjects were more relaxed and therefore less likely to be hypervigilant. Nominal After High Load was the one Workload Level where there was an active task but a relative reduction in task difficulty. This same treatment suffered an unique increase in performance error. This performance decrement was reflected in only two of eight Sensory Parameters. However, it occurred in seven of eight Perception Parameters.

Design of experiment allowed comparison of the same Nominal task following workload increase and workload decrease. No interaction occurred for either performance or perception parameters to indicate different strategies while workload was increasing. Subjects responded appropriately to increases in workload by increasing efficiency of their scan strategy in the morning and afternoon. However, when there was a workload reduction in the afternoon, efficiency of the scan strategy dropped relative to changes recorded for the Primary Task/PM/Nominal After Monitor treatments, $p < 0.001$. Subjects scan strategies got lazy and performance reflected this change. Subjects had the

attention assets available to handle the Nominal workload, but chose not to employ the assets.

Of the three potential modelling parameters for perception, Transition Matrix Useful did not possess the three factor interaction. Significant factors and interactions of Percent Matrix Repeat accounted for only 8.0% of ANOVA variance. Only the High Load treatment was significantly different, $p < 0.01$, from the three other factor levels.

Significant factors and interactions of Dual Fixation Gate accounted for the largest amount of ANOVA variance (32.3%) among all Perception Parameters. Since both fixation trapping techniques were activated together for this parameter it is a measure of the deliberateness of the scan pattern. Long fixations, fixations ended by blinks, and fixations from tracking of slow moving symbols would not qualify as Dual Gate Fixations. The parameter increased as Workload Level increased but only the High Load treatment was significantly different from all other factor levels ($p<0.001$). Dual Fixation Gate was the best candidate to model workload/performance among Perception (Strategy) Parameters.

7.8.4. Selection for Performance Modeling - Cognitive Activity

Increased cognitive processing has been linked to fixation duration (Just and Carpenter, 1976; Harris and Glover, 1985) and to increases in Index of Engagement (Prinzel III et al, 1995b). The four Cognitive Parameters were, Index of Engagement, Fixation Time, Fixation Time Change, and Long Fixations. However, only Fixation Time and Long Fixations demonstrated a consistent correlation with Workload Level.

Index of Engagement is an unproven measurement in the operational aviation environment. However, it has proven to be a useful measure of engagement for closed loop tracking tasks. For this study there were consistencies in the index associated with the final approach task, but the same trend did not occur among all subjects or within treatment replicates for a given subject. Index of Engagement patterns did not develop for tasks other than approach. The total range for Workload Levels was 0.713 to 0.792 but the standard deviation for the variable was 0.295. No discernable trends were seen.

Fixation Time Change did not correlate to workload, performance, or any other discernable factor. The one second limit on fixation time may have created an artificial ceiling for this variable at the High Load treatment. However, both Fixation Time and Long Fixations correlated to workload and performance data. The form of the Workload Level interactions was also similar to performance results for both variables. Long Fixations' significant factors and interactions accounted for slightly more ANOVA variance (15.6%) than did those of Fixation Time (14.1%).

Results for both Fixation Time and Long Fixations bore a striking resemblance to portions of Performance Error Rating data. Of particular interest in Figures 7.21 and 7.22 are the PM treatments. The point of greatest interest was the Nominal After High Load point in the afternoon. This point was above the High Load point for both performance and Long Fixations.

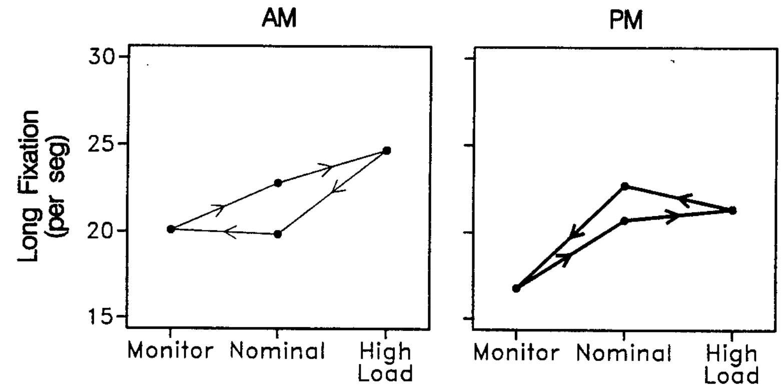

Figure 7.21. Long Fixations interaction of Time of Day/Workload Level

Another interesting similarity between the two parameters was the reversal of flow (counter-clockwise versus clockwise) in the afternoon versus the morning. The patterns of significance were also the same for the parameters in the afternoon. Nominal treatments were the same as the High Load and significantly greater than the Monitor treatment.

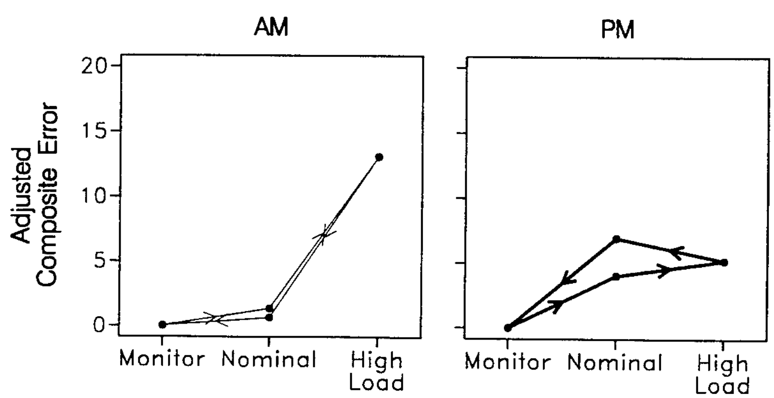

Figure 7.22. Composite Error Rating interaction of Time of Day/Workload Level

Long Fixations is the preferred cognitive modeling parameter due to its marginal advantage over Fixation Time in factor significance level and variance accounted for in the ANOVA model.

Chapter 8.

PERFORMANCE AND PSYCHOPHYSIOLOGICAL DATA RELATIONSHIPS

Workload related treatments accounted for 57.5% of the variance for the Performance Rating non-parametric variable. Six psychophysiological variables referenced at the end of Chapter 7 accounted for 15.5 – 65.9% of ANOVA variance. All of these parameters had significant Workload or Workload interactions. This chapter removes Workload as the mediator to provide direct correlation of performance and psychophysiological parameters through ANOVA. The factors for the ANOVA were:

Performance Level (4)

Group(4)

Subject(Group) (16)

8.1. Performance Level.

The last parameter presented from Chapter 6, Performance Error Rating, was based on operationally accepted limits in aviation. The levels for the non-parametric variable are presented in Table 8.1. The corresponding Composite Error range is shown to the right. The ATC Composite Error limit on any single control axis was two. Exceeding the limit of two on one control axis would result in a Federal Aviation Regulation "Violation." This can result in loss of aviation rating. Performance Error Levels of Monitor, Low, Medium, and High, are derived from Performance Error Rating.

Table 8.1. Factor Levels for Performance Error Rating.

Factor Level	Upper Limit of ATC Performance Error Criteria (Three Axis Total)	Composite Error
0	All Control Axis Less Than Half Of ATC Limit	0-1.9
1	Control Axis Within ATC Limit	2.0-5.9
2	ATC Limits Exceeded But NOT Dangerous	6.0-23.9
3	Performance Error Exceeding Danger Limit	24.0-

The study was designed to provide a similar performance range for all four groups of subjects. However, the Workload levels for each subject were set prior to the simulation start. Subject performance was not strictly controlled to produce specific performance levels because of the artificiality introduced by such controls. Table 8.2 provides breakdown of Group and subject performance by Performance Error Rating Category.

A regression approach was initially considered to describe the relationship between psychophysiological parameters and performance. The regression approach was not presented for two reasons. First, none of the transformations attempted could mitigate heteroscedasticity of the performance data and retain a significant correlation to the parameters. Second, and more importantly, the ATC related levels into which performance was grouped were more operationally pertinent. Operational relevance was a significant concern.

The small number of samples in Rating 3 required the category be collapsed into Rating 2 if the ANOVA was to be accomplished using Group and Subject(Group) factors. Data from the five subjects with more than one replicate for treatment Rating 3 were used to determine if there was a significant difference within Subject between Rating 2 and Rating 3. All six psychophysiological parameters were tested individually for the five

subjects using Tukey's Test. Results show there was no difference between the subjects' psychophysiological data for Rating 2 and Rating 3. Data for Rating 2 and Rating 3 were collapsed into a single factor level for the ANOVA.

Table 8.2. Group and Subject Breakdown by Performance Error Rating

Group	Subject	Monitor	Rating 0	Rating 1	Rating 2	Rating 3
1	4	16	23	13	12	0
1	5	16	21	12	15	0
1	8	16	15	20	11	1
1	14	16	13	20	12	3
2	2	16	24	11	12	1
2	3	16	22	16	9	0
2	9	16	13	16	12	0
2	18	16	20	15	11	2
3	10	16	15	12	20	1
3	13	16	23	14	13	1
3	16	15	23	9	13	1
3	17	16	21	15	9	3
4	6	16	22	16	10	1
4	11	16	23	14	10	1
4	12	16	22	14	12	0
4	15	16	13	16	17	2
Total	16	252	316	236	193	18

There were 568 data points with low performance error. However, 252 of those points were recorded while the subjects were monitoring the simulation. The subjects manually flew 316 of 568 segments in which the performance error level was Low. Performance error for the Monitor level was also low since the simulation was controlled by either the autopilot or an expert aviator. Since it was possible that subjects might react differently to Low Performance Error when they were actually flying, versus when the simulation was controlled from another source, the Monitor level was retained for this

ANOVA. The Low factor level included only the manually flown segments meeting the Rating 0 criteria in Table 8.1.

The second factor level, Medium, was all hand flown. Performance error reached a point where some focused effort was required to ensure the simulation did not exceed ATC limits. All subjects were reminded of the gravity of exceeding these limits during the pre-brief. Of the 1024 data segments, 236 segments resulted in Medium performance error.

Finally, the last two Performance Error Ratings were combined into the High factor level. Operationally, there is a difference between the gravity of exceeding ATC limits (Rating 2) and the episodes of life threatening performance error composing Rating 3. However, the higher level of performance error was undoubtedly mitigated by the knowledge that subjects could not actually crash the simulator and kill themselves.

8.2. ANOVA for six Psychophysiological Parameters.

A three factor ANOVA (Group, Subject(Group), and Performance Level) was performed for six psychophysiological parameters, Pupil Diameter Change, Peripheral Temperature Change, Saccade Time, Short Fixation, Dual Fixation Gate, and Long Fixation. Factors of Performance Level, Group, and Subject(Group) were considered. Performance Level was significant for all six parameters which constituted the best subset of psychophysiological parameters for modeling. The Group factor was significant for only Dual Fixation Gates, $F(3,12) = 3.63$, $p < 0.05$. The Group/Performance interaction was never significant. Results are summarized in Table

8.3. In the following series of figures, whiskers represent the minimum significant difference from Tukey's Test.

Table 8.3. Summary of ANOVA for Six Psychophysiological Parameters.

Dependent Variable	DF	SS Effect	SS Error	F-value	P-value	G-G P-value	ω^2
Pupil Diameter Chg	3,36	356.40	482.26	8.87	0.0002	0.0005	0.347
Peripheral Temp Chg	3,36	114.20	20.570	66.62	0.0001	0.0001	0.768
Saccade Time	3,36	4.5E-5	1.5E-5	9.13	0.0001	0.0009	0.058
Short Fixation	3,36	180.49	305.07	7.10	0.0007	0.0034	0.098
Dual Fixation Gate	3,36	0.025	0.039	7.61	0.0005	0.0017	0.075
Long Fixation	3,36	238.63	220.21	13.00	0.0001	0.0001	0.110

F-tests for Performance Error (Perf) used Perf*Subject(Group) as the error term.

8.2.1. Performance and Pupil Diameter Change.

Pupil Diameter Change was an Arousal Parameter which showed significant correlation, $F(3,36) = 8.87$, $p < 0.001$, with the level of performance error. The parameter accounted for 34.7% of ANOVA variance. Pupil Diameter grew smaller with the Monitor treatment and larger as the level of performance error increased (Fig. 8.1). The amount of change was the same for both the Medium and High error treatments, indicating subjects became incrementally more aroused as their performance worsened. It is notable that this parameter differentiated among Monitoring and active performance levels.

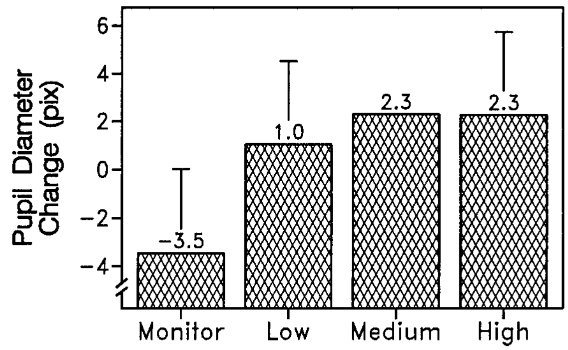

Figure 8.1. Pupil Diameter Change versus Performance Level

8.2.2. Peripheral Temperature Change and Performance Level.

Peripheral Temperature Change was a second Arousal Parameter accounting for an even higher portion (76.8%) of ANOVA variance, $F(3,36) = 76.62$, $p < 0.001$. Figure 8.2 illustrates the similarity between the two Arousal Parameters. Again, the Medium and High treatments showed an incremental decrease in Peripheral Temperature and differentiated among the Monitor and active Performance Levels.

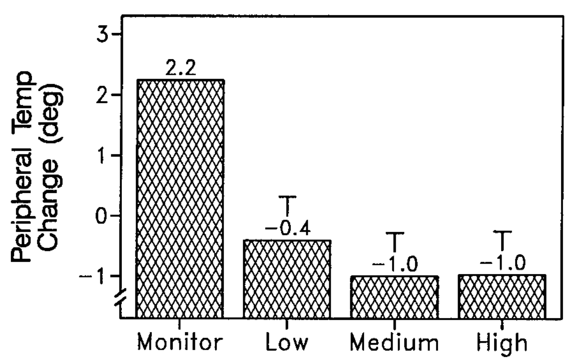

Figure 8.2. Peripheral Temperature Change versus Performance Level

8.2.3. Saccade Time and Performance Level.

The Saccade Time parameter accounted for only 5.8% of ANOVA variance but this Early Perception Parameter was complimentary to arousal parameters. While the arousal parameters differentiated among Monitor and active Performance Levels, Saccade Time provided resolution between the High Performance Error treatment and the other three levels, $F(3,36) = 9.13$, $p < 0.001$. The High level is particularly important because it contains segments in which the performance was poor enough to warrant disqualifying a pilot from duty. Figure 8.3 illustrates the difference between the High treatment compared to the Low and Medium treatments, $p < 0.05$.

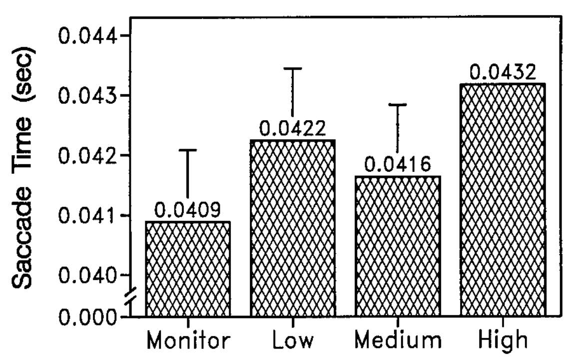

Figure 8.3. Saccade Time versus Performance Level

8.2.4. Short Fixation and Performance Level.

The Short Fixation Parameter was unique because it accounted for more Performance Level variance (9.8%), than it did Workload variance. This Perception

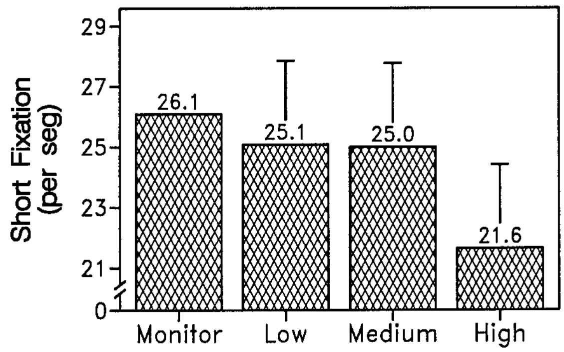

Figure 8.4. Short Fixation versus Performance Error Level

Strategy parameter was developed in hopes of determining the level of automaticity employed in the scan pattern. Although data resolution was not sufficient to determine the existence of automaticity-as-memory, it did provide another means to differentiate between the High Performance Error treatment and the other treatments, $F(3,36) = 7.10$, $p < 0.001$. However, this parameter did not provide resolution on the Monitor treatment.

8.2.5. Dual Fixation Gate and Performance Level.

This second Perception Strategy parameter was expected to differ from the Short Fixation parameter since this parameter would be indicative of the overall visual scanning strategy, while the Short Fixation parameter was designed to be indicative of a particular part of the Perception Strategy (automaticity). Figure 8.5 illustrates how Dual Fixation Gate again differentiated the High Performance Error treatment, $F(3,36) = 7.61$, $p < 0.001$, however, significance for Greenhouse-Geisser was $p < 0.01$ due to heteroscedasticity. It was the only parameter displaying a significant difference for Group, $F(3,12) =$, $p < 0.05$ (0.045). The Group significance was due to a drop in the

fraction of Dual Fixation Gates for the Novice group compared to all other groups. The strategy variables were the most likely candidates for a significant Group effect.

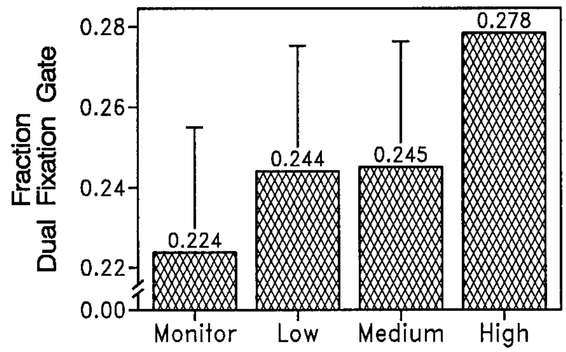

Figure 8.5. Dual Fixation Gate versus Performance Level

8.2.6. Long Fixation and Performance Level.

Although Long Fixation did not account for the largest percentage of ANOVA variance (11.0%) among the psychophysiological variables, it was the most sensitive. The only Performance Levels Long Fixation did not differentiate between were the Low and Medium levels, $F(3,36) = 13.0$, $p < 0.001$. The Low and Medium treatment levels both represented nominal manual performance. These nominal treatments were higher than the Monitor level, $p < 0.001$, and lower than the High treatment, $p < 0.001$. Figure 8.6 illustrates the uniqueness of this Cognitive Processing variable.

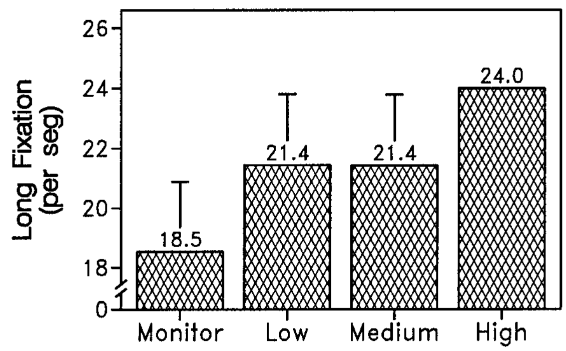

Figure 8.6. Long Fixations versus Performance Level

8.3. Summary of Psychophysiological Parameter Correlation to Performance Level.

The six psychophysiological parameters showing greatest promise in Chapter 7 were analyzed to determine if they were correlated to aviation performance. All six parameters displayed significant differences between at least two performance levels. However, all six parameters do not appear to be unique since there were only five profiles for the factors.

The Arousal Parameters were alike in level of significance and profile. Both Peripheral Temperature Change and Pupil Diameter Change were significantly different for the Monitor Performance Level when compared to any of the manual performance levels. At the Low level both trended toward the Monitor level but were not different from the Medium or High levels. These parameters accounted for the largest portions of ANOVA variance among the six variables, but Pupil Diameter Change accounted for a smaller portion due to greater variability in the parameter.

The Early Perception parameter, Saccade Time, was the only parameter that showed an inversion for the Low and Medium treatments (i.e. – The data values formed a slight zigzag through the four levels.) This difference between the Low and Medium

treatments was not significant, but both middle levels were not different form either the High or Monitor Levels.

One Perception Strategy variable was expected to indicate automaticity. Short Fixation, was unique in that the High Performance Level treatment was the only different treatment. The High error treatment is the most important treatment due to the grave consequences associated with it.

A second Perception Strategy parameter, Dual Fixation Gate, showed the increase in deliberateness that occurred as subjects attempted to correct performance error. Again the two middle levels, Low and Medium, were the same, but they were both different from the Monitor and High treatments. In addition, the Dual Fixation Gate parameter showed a significant Group effect for the Novice group which would be the one group expected to display a different (lack of) strategy.

Finally, the sole Cognitive Parameter, Long Fixations, was similar in form to Dual Fixation Gate. The two middle treatments were alike but unlike Dual Fixation Gate, they were both different from the Monitor and High treatments. In addition, there was no Group significance. None of the six parameters exhibited a significant difference between the Low and Medium treatments indicating that psychophysiological parameters are not very reactive within normal operating ranges.

Chapter 9.

DISCUSSION

The first objective of this study was to determine the relationship between subjective workload levels and operational performance limits. The second objective was to associate 29 psychophysiological parameters with the same subjective aviation workloads to determine if the parameters varied significantly with workload. The final objective directly associated psychophysiological parameters with performance patterns to accomplish the primary goal of the study. This goal was to determine if psychophysiological parameters indicate attention related problems associated with increases in performance error. All three objectives and the study goal were accomplished.

Operational aviation performance levels were closely related to workload level indicating appropriate assumptions were used in design of experiment. In addition, 19 psychophysiological parameters varied predictably with workload level. Six of the nineteen psychophysiological parameters described earlier contributed uniquely to describing different areas of performance decrement.

9.1. Aviation Performance Measures

Developing performance measures related to operational conditions yet distributed in a manner to allow statistical analysis was a significant challenge. Chapter 6 presented performance results and outlined a procedure to produce operationally related metrics. Commercial and military pilots are not concerned with metrics like Root Mean Squared (RMS) error, furthermore no transformation exists relating RMS error to

operational metrics like ATC violations and crashes. The metric, Performance Error Rating, discussed below provided a link to safety and ATC limits.

9.1.1. Performance Error Rating Development

Airspeed was not a controlled variable. Like the two other performance error measures, airspeed error varied appropriately with workload. Significant factors and interactions accounted for 36% of ANOVA variance. However, using this raw performance metric presented three serious problems. First, airspeed error was only one of three performance measures, therefore performance tradeoffs among control axis would be lost. Second, raw data was not related to operational limitations. Third, data displayed significant heteroscedasticity. An operationally relevant model must include all three axes of error and permit the product of the three axes to submit to statistical analysis.

Processing airspeed error as a function of operational airspeed limits increased the amount of variance accounted for to 42%. This adjustment accounted for the more stringent airspeed control required on final approach. In addition, the adjusted index allowed addition of the other two adjusted performance error indices, Adjusted Cross Track Error and Adjusted Vertical Error. The three indices added together comprised the Composite Performance Error Index which provided a complete picture of performance error and control axes tradeoffs.

Error levels among the three indices varied as subjects reprioritized the control axes while searching for an optimal control strategy. Differences varied by subject. When the indices were added to form the Composite Performance Error Index, Group

was not a Significant factor and Subject(Group) was only marginally significant. Comparison of the Workload Level/Time of Day interactions across the three adjusted indices demonstrated the different priorities subjects placed on controlling error for the different axis. Adding the three indices together pooled the three sources of error to reduce variation due to differences in subject strategies. Furthermore, significant factors and interactions in the Composite Error Index rose to 63.2% of ANOVA variance. However, variance among factor levels was still not homogeneous, $p < 0.001$.

Several options were considered to deal with the heteroscedasticity. Numerous logarithmic transformations were studied, however, one of two outcomes resulted from all transformation attempts. Either heteroscedasticity remained, or the ANOVA variance accounted for by significant factors dropped by more than 20%. The best solution found was transformation into the non-parametric, Performance Error Rating, related to the operational limitations briefed to the subjects. Composite Performance Error was converted into the non-parametric, Performance Error Rating, by categorizing performance error levels relative to ATC limits.

Three error levels were selected to represent performance error while subjects manually controlled the simulation. It was reasoned that subjects would exhibit different levels of stress and different scan patterns depending on their proximity to ATC limits. Level One and Level Two were both within ATC limits but Level One was low error, while Level Two was close to the limit at which an aviator would risk de-certification. The other two levels were at the extremes of no error while the subjects were monitoring the simulation, and error large enough to warrant loss of aviation rating. Incidents in which subjects produced dangerously high levels of performance error were included in

the latter factor level.

The assumption that subjects would exhibit differences in psychophysiological parameters for Level One and Level Two was incorrect. There was no set of psychophysiological parameters to differentiate between these two nominal factor levels. It did not matter if the subject was close to no error, or close to the ATC limit. Nominal performance resulted in one level of response for all psychophysiological parameters. However, there were significant differences among all other factor level combinations. Some psychophysiological parameters differentiated between Level Zero (Monitoring) and the three higher factor levels, others between Level Three and the three lower factor levels. Finally, one parameter differentiated among three Performance Error Rating levels, Level Zero, Levels Two and Three, and Level Three.

Once data was converted into Performance Error Rating, variance was homogeneous among factor levels. Heteroscedasticity among ANOVA factor levels for the performance data had been due to 15 large performance deviations. These deviations spanned a large range of performance error but all of these deviations were the same by aviation standards. They would have resulted in revocation of aviation rating. These deviations collapsed into a single Performance Error Rating, Category Three. ANOVA variance accounted for by significant factors and interactions increased from 63.2% for Composite Performance Error to 73.4% for Performance Error Rating.

In addition, creation of a qualitative index exposed an important qualitative trend in the Time of Day factor levels which was not previously seen. The composite error in morning and afternoon was not significantly different because the morning had 15 large performance deviations which could have resulted in an accident, while the afternoon had

128 lesser deviations (ATC violations) which roughly equaled total error from the morning deviations.

Creation of Performance Error Ratings accomplished three important milestones in completing the first objective of this study. First, the ratings related performance error to operational limits understood and accepted by aviators. Second, Performance Error Ratings created a normal data population with uniform variance for the manually flown segments. Finally, the Performance Error Rating accomplished the first objective of the study. It demonstrated the relationship between Workload levels and performance error and provided performance factor levels against which psychophysiological parameters could be compared.

9.1.2. Use of Performance Error Rating.

Initially, performance results were divided into three categories, low error, nominal error, and high error. Monitor segments were not used in development of Performance Error Rating since the results were predefined to be zero error. However, background information pertinent to vigilance decrement (Akerstedt, Torsvall, and Gillberg, 1987; Cole and Hughes, 1990) indicated the monitoring of tasks (versus manually performing tasks) could affect the scan patterns. Hence, psychophysiological parameters could also be affected. Therefore, Monitor was added as a fourth level when ANOVA were conducted to compare psychophysiological parameters directly to performance. The psychophysiological results below reconfirmed previous findings by indicating visual scan patterns were significantly different from comparable Monitor performance results. Unfortunately, experimental design allowed only one error level

during the monitoring treatments. Future studies should incorporate different levels of error during monitoring treatments to allow comparison of psychophysiological parameters between all levels of performance error.

9.2. Psychophysiological Measures.

Twenty six eye movement parameters, two peripheral temperature parameters, and an EEG parameter were measured and compared to workload. Two unanticipated interactions affected a number of parameters. The first interaction was related to segment type, and the second involved a reversal in the expected effects of adding a secondary task. These two interactions are discussed first.

9.2.1. Psychophysiological Parameters Related to Segment Type.

Two segment profiles were flown as part of the simulation. The two segment types were approach segment and downwind segment. In the morning the Monitor and High Load treatments occurred on approach segments while the Nominal treatments occurred on downwind segments. In the afternoon the treatments were reversed placing the Monitor and High Load treatments on downwind segments and the Nominal treatments on approach segments.

Seven eye movement parameters displayed a Workload Level/Time of Day interaction that was confounded by the segment type (approach segment or downwind segment). The Monitor and High Load treatments from the morning simulation and the two Nominal treatments for the afternoon simulation were similar; these data were

recorded during tasks accomplished on approach segments of the simulation. Likewise, the four treatments resulting from downwind segments were also similar.

Figure 9.1 illustrates the interaction due to task dependency using the Maximum Ellipticity Parameter. The inverted "V" and "V" shapes appeared for the seven parameters previously mentioned because data were aligned by segment type versus time of day. Arrows indicate the progression of data points (Monitor – Nominal – High Load versus High Load – Nominal – Monitor).

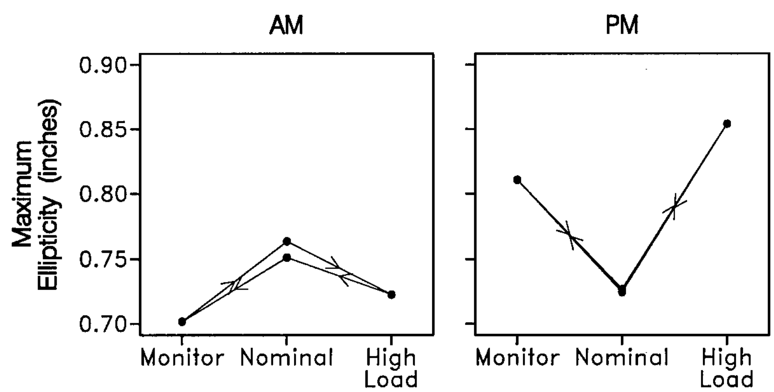

Figure 9.1. Maximum Ellipticity interaction of Time of Day/Workload Level

There was no significant difference among approach segments (four lowest values, 0.70 – 0.74); however, all three downwind workloads resulted in higher, and significantly different values. Video tape review indicated the three dimensional approach tasks differed from the two dimensional tasks in the type and frequency of display scanned. The two dimensional downwind task showed a marked increase in use of the two dimensional Horizontal Situation Display (HSD; Top-down View). Interestingly, the afternoon Monitor and High Load treatments both had a higher

ellipticity than the morning Nominal treatments. Similar results occurred when subjects worked a consuming two dimensional task and when distracted by a moving two dimensional while monitoring performance. Maximum Ellipticity results reflected task type regardless of workload.

The seven parameters that could be associated with segment interaction were Angle Gate Fixations, Velocity Gate Fixations, Fixation Size, Fixation Size Change, Maximum Ellipticity, Maximum Ellipticity Change, and Percent Transition Matrix Symmetric. These parameters are particularly interesting because they hold the potential to identify the task on which the operator is engaged. This information would be useful to Human Factors applications for task decomposition and adaptive display design. However, the close alignment with segment type precludes further analysis for the purposes of this study.

9.2.2. Secondary Task Effect on Performance and Psychophysiological Parameters.

For two reasons, this study was designed with a secondary task on each segment. First, operational administrative tasks are a normal intrusion on the primary instrument cross check. Second, it was believed the addition of a secondary task would increase the potential for performance error due to task overload. Performance error was necessary to the goals of the study since the ultimate goal was to demonstrate the relationship between performance error and psychophysiological parameters.

Surprisingly, performance error was consistently lower in the presence of a secondary task. A review of experiment videos tape revealed subjects using the flexibility built into each segment to minimize error before attempting the secondary task.

This strategy was much like that of a car driver consuming fast food after going to a drive through window. Normally, the driver waits until clear of poles, buildings, and sharp turns before attempting the secondary task of opening and eating food. Subjects knew they had to complete the secondary task while on the required segment, but they waited until they believed they had minimized error before attempting the secondary task.

The lower level of performance error in the presence of a secondary task was reflected throughout performance results. The significance of the interaction was evident in the Secondary Task/Workload Level interaction. Two implications follow from this interaction. First, error in the presence of the secondary task was lower than without the extra task. If performance had been level or worse with the secondary task, the implication would be that the two tasks were competing for similar attention resources (Wickens, 1992) and as a result the primary task suffered in performance.

If tasks were competing for attention resources, performance should not have improved. At best, attention would have been reallocated to keep performance stable while accomplishing the secondary task with the attention resources remaining. However, performance improved in the presence of the secondary task indicating the extra task had no significant effect on performance. Therefore, the secondary task used attention resources that did not compete with the primary task. This hypothesis was reinforced by the significant increase in short fixations for segments with a secondary task. Short fixations indicated an alternative processing method being utilized to complete the secondary task.

Second, since error was minimized, there was a more smooth predictable progression of performance error through the Workload Level treatments with a

secondary task. Excess performance error and the associated variance was pooled in the Primary Task treatments. Levels of performance error were more predictable with the addition of the secondary task and variance for these treatments was one-third that of the primary task treatments. Use of the "Fast Food Effect" can significantly increase the quality of performance data.

9.2.3. Experimental Considerations

Thirteen out of 29 parameters varied predictably with Workload. A majority of the parameters also exhibited interactions which correlated to performance decrements. Clearly, the second objective of the study, "Correlation of psychophysiological parameters to Workload Levels," was accomplished. However, three problems surfaced with this accomplishment.

First, design of experiment was incomplete because there was no provision to allow for performance error while subjects were monitoring the simulation. Because of this oversight, the very dangerous situations resulting from inattention while monitoring, like the Cali, Columbia airline crash, were not captured. This issue could have been overcome by allowing the copilot to fly poorly or misinterpret ATC instructions. Alternatively, the autopilot could have been purposefully misprogrammed as occurred in the Cali accident (Simmon, 1996). Unfortunately, study setup time precluded the use of either option.

Second, variance in psychophysiological data was significant enough to preclude use of a regression approach for most parameters that varied significantly with Workload. The percent of ANOVA variance accounted for by significant factors and interactions

ranged from 4.4% to 66.2%. Parameters accounting for a low percentage of Workload variance displayed greater heteroscedasticity. Only five of the thirteen parameters varying with workload possessed uniform variance. These five parameters all accounted for more than 15% of ANOVA variance. One additional parameter, Peripheral Temperature Change, was slightly heteroscedastic but accounted for 66.2% of performance variance. Of the six parameters, two were related to fixation duration (Short Fixation and Long Fixation), two were related to arousal (Peripheral Temperature Change and Pupil Diameter Change), the fifth was Saccade Time, and the last was Dual Fixation Gate. Dual Fixation Gate tracked the percentage of fixation identified by both saccadic velocity and angular movement from the fixation location. When these six parameters were analyzed via multiple regression, all six were significant. Results were not presented here due to the heteroscedastic nature of the performance data.

The third problem was that the cycle parameters showed significant trends, but a number of treatment cells had no data. This prevented analysis of interactions and variance within subjects. Review of subjects for which this was a problem, showed it was due to subjects' flexible approach to anchoring their scan cycles. It was assumed the most frequently visited area of interest would be a suitable anchor for cycle analysis. However this was not the case as evidenced by the incomplete data. In fact, it appears the anchor for viewing cycles depended on the source of performance error. For example, if the perceived performance problem was altitude control, the subject anchored the cycle on the altimeter, but if the problem was airspeed control the cycle was anchored on the airspeed indicator. A new anchor schema based on a situated cognition approach (Zhang and Norman, 1994) might mitigate this problem.

There were three useful findings from the study. First, the quality of the eye movement data produced by the analysis process was excellent. Fixation and saccade times corresponded well with established norms. These basic parameters varied as expected with workload, validating the experimental design and analysis code. Concern about sufficient resolution to make saccade time precise enough for analysis was unwarranted.

The second useful outcome was the outstanding correlation of Peripheral Temperature Change to Pupil Diameter Change; this is a known albeit highly unstable arousal metric. The correlation indicated Peripheral Temperature Change was a slower responding, but more stable measure of arousal. In addition, Peripheral Temperature Change accounted for 66.2% of ANOVA variance in Performance Error Level and produced the same significant interactions found in performance variables. These interactions occurred with Time of Day and Primary/Secondary Task. Changes in pupil diameter and peripheral temperature both indicated a drop in arousal when comparing Nominal tasks to the High Load tasks they followed.

Finally, within the three factor interaction associated with the vigilance decrement, the Percent Transition Matrix Symmetric increased with decreasing workload. This result indicated a tendency for free, undirected, scanning (Ellis, 1986). However, Peripheral Temperature remained stable indicating subjects remained in a high state of arousal. When workload decreased from High Load to Nominal, the visual scan pattern changed inappropriately to seek new visual stimulus when the normal working patterns should have been maintained to accommodate a nominal workload. This inappropriate change correlated to an increase in performance error.

Did the change in visual scan pattern precede, coincide with, or follow the increased performance error? A time correlation between specific performance decrement incidents and the Transition Matrix Symmetry parameter may answer this question, but the current analysis approach does not provide sufficient temporal correlation. However, changes in Transition Matrix Symmetry and fixation duration parameters did confirm the change in visual scanning strategy hypothesized in Chapter 2. Psychophysiological Parameters were successfully correlated to workload, accomplishing the second objective of the study. Discussion of the six most significant parameters follows.

9.2.4. Psychophysiological Parameters Related to Arousal

Level of arousal can have an effect on all phases of information processing. Since these were global parameters, it was not surprising that the two arousal parameters, Peripheral temperature Change and Pupil Diameter Change, were most closely related to performance. In fact, arousal parameters were more closely related to performance than was workload.

Pupil Diameter Change. Pupil diameter decreased for Monitor treatments but increased for all manual treatments regardless of performance level. This parameter changed rapidly and predictably for all factors and interactions except in the afternoon in the presence of a secondary task. An unexpected result occurred in the afternoon with a primary task treatments. These treatments yielded performance decrements that were not mirrored by changes in pupil diameter. Subject pupil diameter changed predictably with workload on these treatments, not with performance. This fact was important when

considering the relationships among Pupil Diameter Change, Peripheral Temperature Change, and performance.

Peripheral Temperature Change. Stick control inputs were reviewed as related to changes in peripheral temperature. Statistical analysis showed a strong correlation between stick manipulation and decrease in peripheral temperature. This finding would support control of peripheral temperature by the stress mechanism since peripheral temperature was decreasing with increased workload. In addition, review of data from sixteen subjects showed periods of time exceeding 36 seconds where changes in the stick manipulation patterns had no corresponding change in temperature. Likewise, there were periods in which stick manipulation patterns did not change but peripheral temperature changed rapidly. Peripheral Temperature did not always mirror workload.

Indeed, the high correlation of stick movements with peripheral temperature could indicate peripheral temperature was a low level function related to physical exertion. However, it became evident that temperature changes were being driven by higher level processes when considering anecdotal situations. One particular commercial airline pilot was convinced the simulation would include multiple tests of his ability to handle emergency procedures. Throughout the morning simulation he stood at the ready for these emergencies, and his peripheral temperature did not vary by more than two degrees. Thirty minutes into the afternoon simulation his peripheral temperature was still a flat line until he turned to the first officer and said, "We really aren't going to do emergency procedures, are we?" Within sixty seconds hearing, "That's right!" his peripheral temperature rose six degrees. Another subject's temperature went down at the same time she complained of a headache. Several similar episodes occurred with this and other

subjects. These changes did correlate with changes in stress in addition to changes in task difficulty. Although arousal was closely related to physical activity (simulator control manipulation), additional factors related to arousal, such as anxiety, pain, and comfort level also affected the parameter.

Peripheral Temperature Change mirrored both workload and performance better than any other parameter. Like Pupil Diameter Change, this parameter indicated an increase in arousal (with decreased temperature) for all manual control segments regardless of performance. The parameter was more closely related to performance than Pupil Diameter Change because it mirrored the afternoon performance decrement while pupil diameter did not. Pupil diameter changed appropriately to reflect workload change, but there was no change in peripheral temperature for these treatments. There are at least three possible explanations for changes in pupil diameter where there were no changes in peripheral temperature.

First, peripheral temperature lagged changes in workload by as much as 60 seconds and in some cases changed slowly even after that time period. Pupil diameter change occurred almost instantaneously with workload changes. The difference could be due to the time lag. However, Peripheral Temperature Change was calculated as the difference between points recorded about six minutes apart. Temperature changes did not lag by that long a period. In addition, only four of sixteen treatments were zero change treatments; this was not a systemic trend. Lag due solely to physiological changes would likely be more predictable and systemic.

Second, thermal inertia could have been due to the time of day. The standard deviation of the peripheral temperature was smaller in the afternoon than the morning.

However, it is more likely the parameter varied less because subjects were more relaxed (average peripheral temperature was three degrees higher in the afternoon than morning) and less likely to react. Afternoon data did not display any systemic trends. The unique results occurred for only one afternoon treatment, Nominal After High Load/Primary Task.

Third, the two arousal parameters represent different parts of arousal which is an outward manifestation of attention. Pupil diameter always reacted to the perception of workload change while peripheral temperature did not. However, peripheral temperature tracked much more closely to performance, regardless of workload. Pupil diameter represented the perception phases of information processing, and peripheral temperature represented the decision/response selection phases of information processing. When these two parameters were synchronous performance was predictable for any given workload level. (i.e. – Nominal workload resulted in good performance and High Load resulted in performance decrement.) However, performance results were not related to workload when the afternoon performance decrement occurred, and the arousal parameters were out of synch. The mismatched arousal parameters provide a signal indicating the Time of Day/Task combination where a performance decrement might be expected, but not the workload level. Other psychophysiological parameters, such as Long Fixation, did provide specific evidence of the performance decrement.

9.2.5. Psychophysiological Parameters Related to Saccade Time

Saccade Time results were similar to those of Fixation Duration parameters with two notable exceptions. First, there was no three-factor interaction for Saccade Time;

fixation duration parameters possessed this interaction. Second, the two-factor interaction of Workload/Time of Day was more uniform and predictable for the Saccade Time parameter.

Since the three-factor interaction was an indicator of a performance decrement not related to workload, these results indicate Saccade Time varied more closely with workload, than with performance. In addition, the predictable progression and greater significance among workload levels makes Saccade Time an appealing candidate for modeling aviation workload level.

9.2.6. Psychophysiological Parameters Related to Fixation Type

Three eye movement parameters were a product of fixation analysis code. The end of a fixation was determined by either onset of saccadic movement or movement outside a preset fixation angular size. Those fixations identified by both the saccadic velocity and angular movement were recorded as part of the parameter Percentage Dual Fixation Gate.

Fixations used to track slow moving display indicator movements were trapped by the angle gate alone. Fixations resulting from temporary, short term loss of the pupil image were trapped by the velocity gate alone. The latter fixations were often due to subjects blinking. Tracking fixations and high frequency of blinks were both indicators of an undisciplined aviation crosscheck. These scan pattern artifacts should be absent from a highly motivated crosscheck. Dual Gate fixations indicate use of a more disciplined crosscheck. This explains the increase of Dual Gate fixations as workload

increased. Again, there was no three-factor interaction which indicated this parameter was more closely related to workload level than performance.

9.2.7. Psychophysiological Parameters Related to Fixation Duration.

Although fixation duration increases with task difficulty (Just and Carpenter, 1976; Harris and Glover, 1985), many different mechanisms for the increase were proposed, but not validated in literature. The time increase could be due to a reduction in the number of short fixations, an increase in the number of long fixations, a uniform increase in the duration of all fixations, or some combination of these factors. In this study, the Long Fixation and Short Fixation parameters were more closely related to performance than was the generic Fixation Duration parameter. The parameters Long Fixation and Short Fixation also provided unique insight into changes in viewing pattern indicated by several interactions.

Short Fixation Parameter. The Short Fixation Parameter was originally conceived as a measure of automaticity. As such, it was different from the other Perception Parameters because it was targeted toward a narrow aspect of perception. Other Perception Parameters were assumed to describe the broader perception process between early perception and response selection. Whereas, the Short Fixation parameter was designed to confirm automaticity-as-memory.

If it were found to exist, a bimodal distribution of fixation times and a corresponding distribution of reaction times would indicate two separate loops of the HIP model were used. Such a finding would warrant modification of the HIP Model to include the automaticity-as-memory loop. However, average fixation times did not

provide sufficient resolution to determine if short fixations and normal fixations were distributed in a bimodal manner. Reanalysis of individual subject data may further explain this theory if the clear presence or absence of a bimodal distribution is confirmed.

Workload Level was not a significant factor for Short Fixation. The lack of significance becomes clear when considering the Secondary Task/Workload Level interaction. The math tasks performed as part of the secondary task were designed to discourage sequences of short fixations on the primary instruments. In reality, the opposite occurred. The number of short fixations increased slightly with workload. This result indicated subjects attempted to use more short fixations to complete math tasks while performing the primary aviation task, although the increase was not significant. The small difference among the treatments corresponded well with the different levels of performance error recorded in the treatments. The number of short fixations decreased with the decreased processing required when confirming nominal conditions.

An increase in the number of short fixations was an indicator of good performance, without regard to workload. This result was not intuitive since high workload is associated with increased fixation duration. In this study, good performance in a high workload treatment was accompanied by a higher frequency of short fixations. This could be an indicator of the automaticity associated with good performance in high workload situations.

The three treatments with a significantly lower frequency of short fixations were also the segments with highest performance error. The number of short fixations decreased significantly when subjects were attempting to problem solve their performance error. The resulting increase in fixation duration was predicted from

literature and Hypothesis One of this study. Absence of short fixations was a predictor of poor performance and should be considered as a candidate parameter to warn of dangerous situations.

Long Fixation Parameter. Long Fixations accounted for the largest percentage of ANOVA variance (15.6%) among the three fixation duration variables. It was also the only psychophysiological parameter with three significantly different levels corresponding to three different levels of Performance Error Rating. This was the best candidate psychophysiological parameter to model aviation performance and warn of dangerous situations.

Performance Error Rating was divided into four factor levels ranging from Monitor to High Error. The frequency of long fixations was lowest when subjects were monitoring simulator performance and highest when there was significant performance error by ATC standards. Significance with respect to ATC standards was important in determining a parameter to prevent aviation accidents because ATC standards were designed with a buffer for safety considerations. Pilots may, "lose their wings," if they are more than 300 feet off of altitude, but the nearest aircraft should be at least 1000 feet away in altitude. This 700 foot buffer creates a performance zone where subjects changed scan patterns to problem-solve before they became dangerous. Identification of unique psychophysiological parameters associated with this high error performance zone accomplishes the goal of this study.

In some circumstances it would be appropriate for the aviator to exhibit a high frequency of long fixations. For example, if a commercial pilot was required to determine an entirely new route of flight after airborne or if a military pilot needed to

devise a new attack strategy after exhausting planned options, the frequency of long fixations should increase. An appropriate baseline for the required task would be important to determine if there was a departure from the norm. Once a nominal baseline was determined, as in this study, a significant change in frequency indicates a processing problem and in 81% of the cases an associated performance problem.

In the case of the Cali, Columbia accident the Captain would have better understood the extent of the First Officer's workload from a significant increase in Long Fixation frequency. The Captain may have taken a more active role in managing the load. More importantly, if the First Officer had known the Captain was confused (even though he would not admit it), the seconds saved could have allowed a climb in time to save the aircraft. When the First Officer said, "Just doesn't look right," one of the last things the Captain said about their strange heading was, "... let's press on..." (Simmon, 1996). A record of the pilot's long fixation frequency may have told enough of a different story to convince the First Officer to abandon the approach, and climb.

Workload in instrument flight is relatively predictable compared to that in tactical military aviation. The high workload coupled with a reduction in manpower available, undoubtedly led to more episodes of problem solving and the associated high frequency of long fixations.

9.3. Percent Useful Transition Matrix Parameter.

Of the three potential modeling parameters for perception, Transition Matrix Useful was perhaps the most intriguing. Transition Matrix Useful produced results more

like the Arousal (Attention) Parameters. There were no two and three factor interactions like those found in most performance and psychophysiological parameters.

The scatter plot showing the Percent Transition Matrix Useful relationship with composite performance error (Fig 9.2) illustrated a bimodal distribution. As expected, the majority of the data points, on the right half of the figure, indicated scan strategy was more efficient as the workload increased. These high workload (overload) points produced significant levels of performance error in some cases, but for the majority of the data points nominal performance was maintained with a trend toward increasing performance error.

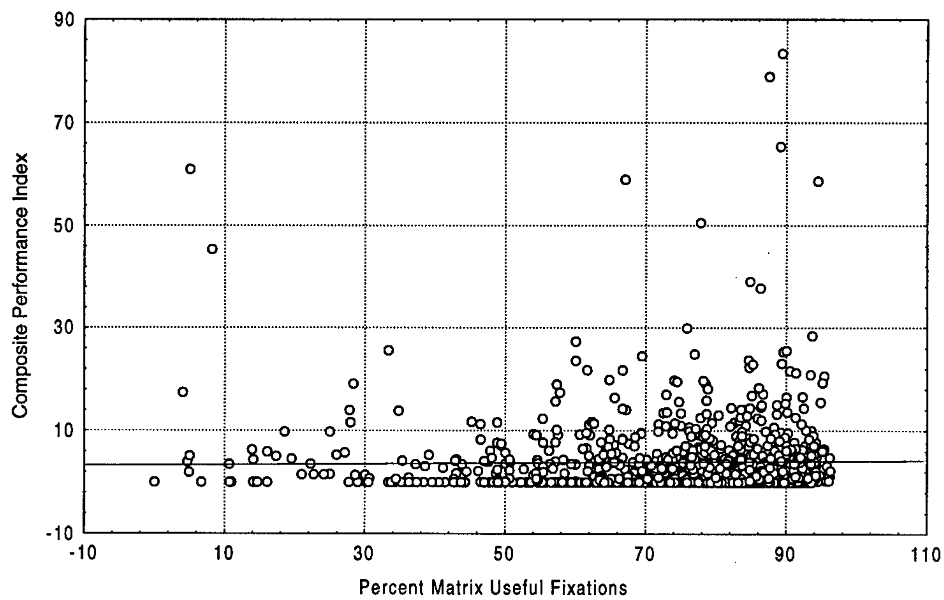

Figure 9.2. Percent Matrix Useful versus Composite Performance Index

However, the interesting portion of the scatter plot was the left side and center of the plot. At approximately 43% Useful Fixations there was a local minimum indicating enough attention to the required indicators for good performance, but also a high

percentage of fixations that were not useful. This large portion of fixations away from useful indicators may constitute an optimum level of flexibility built into the required scan pattern to quickly detect performance degradation.

To the left of this local minimum the number of high performance error incidents increases again. Unfortunately, this side of the scatter plot, which might correspond to a distracted scan pattern, had an insufficient number of data samples to characterize scan patterns in that portion of the plot. As previously discussed, introduction of planned performance error during Monitor workload treatments may have produced clearer trends for analysis .

The U-shaped form of this performance error parameter was also reminiscent of the inverted U-shape curves of previous attention studies. The early performance drop was due to low arousal (task underload), and the late performance drop was due to high arousal (task overload). Since, these studies reported performance, and not performance error, a U-shaped curve would be expected for performance error. Percent Useful Fixations is intuitively appealing as a parameter demonstrating the relationship between arousal and attention using eye movements. Further research is required to confirm this hypothesis.

Chapter 10.

CONCLUSIONS AND RECOMMENDATIONS

Within the context of this study, aviation performance error was best described in terms of Air Traffic Control (ATC) limits. The non-parametric, Performance Error Rating, related well to operational ATC limitations and provided a good basis of statistical analysis. Addition of a separate rating level was appropriate for situations in which subjects were not manually controlling the simulation. There was no variation in performance for the monitoring segments which precluded matching of psychophysiological parameters among the different performance levels. However, psychophysiological parameters differed significantly between the low performance error manual data points, and the low performance error monitoring data points. This study should be expanded to include moderate and high levels of performance error while subjects are monitoring the autopilot and co-pilot.

Since different factors such as workload, attentiveness, and cognitive processing capability can affect performance, different psychophysiological parameters are needed to completely describe performance. Six different psychophysiological parameters used in this study contributed to describing overall performance. Level of arousal was expected to reflect the "level of attention" for perception, processing, and response execution. The best arousal parameters, Peripheral Temperature Change and Pupil Diameter Change, were the best overall parameters in relating to performance, but individually these parameters only reflected a performance decrements related to workload and other stressors. Performance decrements at nominal workload levels were missed.

Workload was described by Saccade Time and Dual Fixation Gate parameters. These parameters reflected the level of efficiency employed by subjects as workload increased. Since these parameters were closely related to workload, they reflected a performance decrement related to workload, but they also missed the decrement unrelated to workload. Despite missing the performance decrement, these two and a seven other parameters show great promise in providing real time feedback on workload levels and the type of task in which operators are engaged.

Elements of cognitive performance were described by the Long Fixation and Short Fixation parameters. A high frequency of Long Fixations (fixations greater than 500 ms) was indicative of problem solving activity. The Long Fixations parameter was the best single parameter in accounting for ATC violations that could lead to danger. A high frequency of Short Fixations was indicative of efficient processing. However, the efficiency was not related to only to workload since subjects used large numbers of short fixations when monitoring the simulation. Further analysis and research should be accomplished to determine if the Short Fixations parameter is related to automaticity.

All six psychophysiological parameters contributed uniquely to modeling performance error by describing different factors affecting aviation performance in a simulated aviation environment. The next step is to apply these finding toward development of a real-time feedback loop providing the operator and team members with workload assessment and appropriate warning of dangerous situations.

References

Akerstedt, T., Torsvall, L., and Gillberg, M. (1987). Sleepiness in shiftwork. A review with emphasis on continuous monitoring of EEG and EOG. Chronobiology International, 4(2): 129-140.

Anderson, J.R. (1993). Problem solving and learning. American Psychologist, 48(1): 35-44.

Arnegard, R. J. (1991). Operator strategies under varying conditions of workload. (NASA Contractor Report 4385). Hampton, VA: Langley Research Center.

Bahill, A.T. and Stark, L. (1975). Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. Experimental Neurology, 48: 95-106.

Blomberg, J., Giacomi, J., Mosher, A., and Swenton-Wall, P. (1993). Ethnographic Field methods and their relation to design. In D. Schuler and A. Namioka (Eds.) Participatory Design: Principles and Practices (pp. 123-155). Hillsdale, NJ: Erlbaum.

Boff, K.R. and Lincoln, J.E. (Eds.) (1988). Saccadic velocity: effect of saccade distance. Engineering Data Compendium, Vol I, (pp. 490-491).

Breitmeyer, B.G. and Braun, D. (1990). Effects of fixation and attention on saccadic reaction time. In R. Groner, G. d'Ydewalle, and R. Parham (Ed.), From Eye to Mind: Information Acquisition in Perception, Search and Reading (pp. 71-79). North Holland: Elsevier Science Publishers B.V.

Cole, B. L. and Hughes P. K. (1990). Drivers don't search: they just notice. In D. Brogan Visual Search (pp. 407-417). New York: Taylor & Francis.

Comstock, J., Jr. (Ed.). (June, 1987). Mental-state examination (NASA Conference (Publication 2504). Proceedings of a workshop sponsored by NASA and Old Dominion University. Williamsburg, VA: NASA Scientific and Technical Division.

Crawford, D.J., Burdette, D.W., and Capron, W.R. (1993). Techniqes used for the analysis of oculometer eye-scanning data obtained from an air traffic control display. (NASA Contractor Resport 191559). Hampton, VA: Langley Research Center.

Credeur, L., Capon, W.R., Lohr, G.W., Crawford, D.J., Tang, D.A., and Rodgers, W.G., Jr. (1993). Final-approach spacing aids (FASA) evaluation for terminal-area, tine-based air traffic control. (NASA Technical Paper 3399).

Hampton, VA: Langley Research Center.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66 (3): 183-201.

Ellis, S. R. (1986). Statistical dependency in visual scanning. Human Factors, 28(4): 421-438.

Ericsson, K.A., and Charness, N. (1994). Expert performance: Its structure and acquisition. American Psychologist, 49, 725-747.

Eriksen, C. W. (1990). Attentional search of the visual field. In D. Brogan (Ed.), Visual Search (pp. 3-19). New York: Taylor & Francis.

Fischer, B. and Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. Neuropsychologica 25(1): 73-83.

Fisk, A.D., Lee, M.D., and Rogers, W.A. (1991). Recombination of automatic processing

Components: The effects of transfer, reversal, and conflict situations. Human Factors, 33, 267-280.

Fitts, P. M. and Jones, R. E. (1961). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. In H.W. Sinaiko Selected Papers on Human Factors in the Design and Use of Control Systems (pp.332-358). New York: Dover Publications, Inc.

Fitts, P. M., Jones, R. E., and Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. Aeronautical Engineering Review, 9: 24-29.

Greenhouse, S.W. and Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Grimsley, D. L. (1994). Digital skin temperature and biofeedback. Perceptual and Motor Skills, 79: 1609-1610.

Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, and J. Thomas, Handbook of Perception and Human Performance, vol.I (pp. 10-1 - 10-112). New York: John Wiley and Sons.

Harris, C. M. (1990). On the reversibility of Markov scanning in free-viewing. In D. Brogan Visual Search (pp.123-134). New York: Taylor & Francis.

Harris, R., Sr. and Glover, B. (1985). Effects of digital altimetry on pilot workload.

(NASA Technical Memorandum 86424). Hampton, VA: Langley Research Center.

Harris, R., Sr., Glover, B.J., Spady, A. Jr. (1986). Analytical techniques of pilot behavior scanning behavior and their application. (NASA) Techical Paper 2525). Hampton, VA: Langley Research Center.

Harris, R., Sr. and Spady,A., Jr. (1985, May). Visual scanning behavior. Unpublished paper presented at the meeting of the National Aerospace and Electronics Conference, Dayton, OH.

Harris, R., Sr., Tole, J.R., Arye, R. E., and Stephens, A. T. (October, 1982). How a new instrument affects pilot's mental workload. Paper presented at the Annual Meeting of the Human Factors Society.

Healy, A.F., Fendrich, D.W., Crutcher, R.J., Wittman, W.T., Gesi, A.T., Ericsson, K.A., and Bourne, L.E., Jr. (1990). The long-term retention of skills.

Hugdahl, K., Fagerstrom, K. O., and Broback, C.G. (1984). Effects of cold and mental stress on finger temperature in vasospastics and normal S's. Behavioral Research Therapy, 22(5): 471-476.

Ilan, A.B. and Miller, J. (1994). A violation of pure insertion: Mental rotation and choice reaction time. Journal of Experimental Psychology: Human Perception and Performance, 20(3): 520-536.

Johnston, D.M. (1967). The relationship of near-vision peripheral acuity and far-vision search perfromance. Human Factors, 9(4): 301-303.

Just, M.A. and Carpenter, P.A. (1976). Eye fixations and cognitive processes. Cognitive Psychology, 8: 441-480.

Kahnemen, D. (1973). Attention and Effort. Englewood Cliffs, NJ: Prentice Hall.

Kantowitz, B.H. (1992). Selecting measures for human factors research. Human Factors, 34(4): 387-398.

Kecklund, G. and Akerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. Ergonomics, 36 (9): 1007-1017.

Klein, G.A., and Hoffman, R.R. (1992). Seeing the invisible: Perceptual-cognitive Aspects of expertise. In M. Rabinowitz (Ed.), Cognitive Science Foundations of Instruction. (pp. 203-226). Hillsdale, NJ: Erlbaum.

Kopp, U. and Liebig, T. (1990). Computer simulation and analysis of pilots scanning behavior during complex aircraft guidance and control tasks. In D. Brogan, Visual Search (pp. 311-318). New York: Taylor & Francis.

Lave, J. (1986). Experiments, tests, jobs and chores: How we learn what we do. In K. Borman and J. Reisman (Eds.), Becoming a Worker (pp. 140-155) Norwood, NJ: Ablex.

Lee, M.D., and Fisk, A.D. (1993). Disruption and maintenance of skilled visual search as a function of degree of consistency. Human Factors, 35, 205-220.

Liu, Y. (1996). Interactions between memory scanning and visual scanning in display monitoring. Ergomonics.

Liu, Y. (1996). Quantitative assessment of effects of visual scanning on concurrent task performance. Ergonomics, 39(3): 382-399.

Liu, Y. (1996). Queueing network modeling of elementary mental processes. Psychological Review, 103(1): 116-136.

Liu, Y. (1996). Queuing network modeling of human performance of concurrent spatial and verbal tasks. IEEE Transactions on Systems, Man and Cybernetics.

Logan, G.D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39(2): 367-386.

Logan, G.D. (1991). Automaticity and memory. In W. E. Hockley & S. Lewandowsky (Eds.), Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock (pp. 347-366). Hillsdale, NJ: Erlbaum.

Makeig, S. and Inlow, M. (1993). Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. Electroencephalography and clinical Neurophysiology, 86: 23-35.

Miyao, M., Hacisalihzade, S., Allen, J. and Stark, L. (1989). Effects of VDT resolution on visual fatigue and readability; an eye movement approach. Ergonomics, 32(6): 603-614.

Moray, N. and Rotenberg, I. (1989). Fault management in process control: eye Movements and action. Ergonomics, 32(11): 1319-1342.

Mourant, R. R. and Rockwell, T. H. (1972). Strategies of visual search by novice and

experienced drivers. Human Factors, 14(4), 325-335.

Nicholson, A.N., Stone, B.M., Wright, N.A., and Belyavin, A.J. (1989). Daytime sleep latencies: relationships with the electroencephalogram and with performance. Journal of Psychophysiology, 3: 387-395.

Navon, D., & Gopher, D. (1979). "On the economy of the human processing systems. Psychological Review, 86, 254-255.

Ogilvie, R.D., McDonagh, D.M., Stone, S.N., and Wilkinson, R.T. (1988). Eye movements and the detection of sleep onset. Psychophysiology, 25(1): 81-91.

Okogbaa, O.G., Shell, R.L., and Filipusic, D. (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal. Applied Ergonomics, 25(6): 355-365.

Pergola, P.E., Kellogg, D.L., Jr., Johnson, J.M., and Kosiba, W.A. (1994). Reflex control of active cutaneous vasodilation by skin temperature in humans. American Journal of Physiology, 266: H1979-H1984.

Pope, A., Comstock, J.R., Jr., Bartolome, D.S., Bogart, E.H., and Burdette, D.W. (1994). Biocybernetic system validates index of operator engagement in automated task. Unpublished manuscript.

Pope, A. T. and Bogart, E.H. (1994). Identification of hazardous awareness states in monitoring environments. (NASA Contractor Report 921136). Hampton, VA: Langley Research Center.

Posner, M.I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32: 3-25.

Prinzel, L.J., III, Hitt, J.M., Scerbo, M.W. and Freeman, F.G. (1995). Feedback contingencies and bio-cybernetic regulation of operator workload. An unpublished paper presented to the meeting of the Human Factors and Ergonomics Society 39th Annual Meeting.

Rayner, K. and Morris, R. K. (1990). Do eye movements reflect higher order processes in reading. In R. Groner, G. d'Ydewalle, and R. Parham (Eds.), From Eye to Mind: Information Acquisition in Perception, Search and Reading (pp. 179-190). North Holland: Elsevier Science Publishers B. V.

Saito, S. (1992). Does fatigue exist in a quantitative measurement of eye movements? Ergonomics, 35(5/6): 607-615.

Saitoh, O. and Okazaki, Y. (1990). Eye movements: a tool of chronometry in visual information processing. In R. Groner, G. d'Ydewalle, and R. Parham (Eds.), From Eye to Mind: Information Acquisition in Perception,Search, and Reading (pp. 23-40). North Holland: Elsvier Science Publishers B. V.

Schulte, A. and Onken, R. (1995). Knowledge engineering in the domain of visual behavior of pilots. In J.M. Findlay et al. (Ed.), Eye Movement Research (pp. 503-514). Elsevier Science B.V.

Simmon, D. (1997). Human factors analysis of AAL 965 accident at Cali, Columbia. Unpublished report to NASA. Hampton, VA: Langley Research Center.

Simonov, P.V. and Frolov, M.V. (1987). Slow oscillations of psychophysiology parameters in human operators during monotony. Aviation, Space, and Environmental Medicine, 58: 132-135.

Simonov, P.V., Frolov, M.V., Evtushenko, V.F., and Sviridov, E.P. (1977). Effect of emotional stress on recognition of visual patterns. Aviation, Space and Environmental Medicine, 48 (9): 856-858.

Spady, A., Jr. (October, 1977). Airline pilot scanning behavior during approaches and landing in a Boeing 737 simulator. Presented at the ABARD 25th Guidance and Control Panel Meeting/Symposium for Low Altitude and Terminal Area Flight.

Spady, A., Jr. (1987). Airline pilot scan patterns during simulated ILS approaches. (NASA Techical Paper 1250). Hampton, VA: Langley Research Center.

Spady, A., Jr., and Harris, Sr. (1983, October). Summary of NASA Langley's Pilot Scan Behavior Research. Unpublished manuscript presented at the meeting of the Society of Aeronautical Engineers Aerospace Congress & Exposition in Long Beach, CA.

Spady, A., Jr., Harris, R., Sr., and Comstock, R. (1983, April). Flight versus simulator scan behavior. Unpublished manuscript presented at the meeting of the Second Symposium on Aviation Psychology .

Stern, J. (1993). The eyes: Reflector of attentional processes. Presented as the third speaker in the 1993 Human Engineering Division, Armstrong Laboratory Colloquium Series: The Human-Computer Interface. (From synopsis by June J. Skelly in "Gateway", volume IV; number 4).

Stern, J., Schroeder, S., and Stoliarov, N. (1994). Blink, saccades, and fixation pauses during vigilance task performance: I. time on task (Report no. DOT/FAA/AM-94/26). Washington, D.C.: Office of Aviation Medicine, FAA.

(NTIS No. TD 4.210:94/26).

Sternberg, S. (1975). Memory scanning. New findings and current controversies. Quarterly Journal of Experimental Psychology, 27: 1-32.

Takenaka, K. and Zaichkowsky, L.B. (1990). Physiological reactivity in acculturation: a study of female Japanese students. Perceptual and Motor Skills, 70: 503-513.

Tyson, P.D. (1987). Task-related stress and EEG alpha biofeedback. Biofeedback and Self-Regulation, 12(2): 105-119.

Unema,P. and Rotting, M. (1990). Differences in eye movements and mental workload between experienced and inexperienced motor-vehicle drivers. In D. Brogan Visual Search (pp. 191-202). New York: Taylor & Francis.

U.S. Air Force. (1996). 2.4 Instrument maneuvering. Instrument Flight Procedures Air Force Manual 11-217,vol.I. Washington, DC: HQ AF/XOO.

U. S. Department of Transportation. (1989). Accidents reported by motor carriers of property 1989. (Publication No. FHWA/MC-90-018). Arlington,VA: The Scientex Corporation.

van Quekelberghe, R. (1995). Strategies for autoregulation of peripheral skin temperature. Perceptual and Motor Skills, 80: 675-686.

Viviani, (1995). Presentation for the European Conference on Eye Movement. Unpublished.

Wachtel, P. H. (1967). Concepts of broad and narrow attention. Psychological Bulletin, 68: 417-419.

Weiner, E. L. (1977). Controlled Flight into terrain accidents: system-induced errors. Human Factors, 19(2): 171-181.

Wickens, C. D. (1992). Engineering Psychology and Human Performance (2nd ed.). New York: HarperCollins.

Wickens, C. D., Bellenkes, A.H., and Kramer, A.F. (1995). Visual scanning expert pilots: the role of attentional flexibility and mental model development (Tech. Rep. No. UIUC-BI-HPP-95-05). Urbana-Champaign: University of Illinois, The Beckman Institute.

Williams, A. and Harris, R., Sr. (1985). Factors affecting dwell times on digital displays. (NASA Technical Memorandum 86406). Hampton, VA: Langley Research Center.

Zhang, J. and Norman, D.A. (1994). Representations in distributed cognitive tasks. Cognitive Science, 18, 87-122.

APPENDIX A

INSTITUTIONAL REVIEW BOARD ACTION

AND CONSENT FORM

National Aeronautics and
Space Administration
Langley Research Center
Hampton, VA 23681-0001

421 August 20, 1996

TO: Crew/Vehicle Integration Branch, FDCD
Attn: 152/Dr. James R. Comstock, Jr.

FROM: 421/Secretary, Institutional Review Board

SUBJECT: Institutional Review Board Action

The Institutional Review Board (IRB) met on July 26 to review your project entitled, "Performance Effects of Awareness Characterized by Effective and Hazardous States (PEACHES)." This memorandum documents the findings of that meeting. Techniques to be used in the PEACHES project include eye-movement monitoring through a headband-mounted camera system and EEG through an electrode cap worn on the head.

Several actions were required by the IRB prior to starting the tests. The required actions are as follows: the IRB requires that a member of the Board participate as a test subject prior to the use of human subjects in any research, a copy of the study on eye hazards which forms the basis for the assumptions on safety for this program is to be provided to the Secretary of the IRB; and the informed consent form to be used by project personnel must be reviewed and accepted by the Office of Chief Counsel (OCC).

As of the date of this memorandum, each of the actions stated above has been complied with as follows: Mr. Rob Rivers, a member of the IRB, has used both the head-mounted camera system and the EEG electrode cap; a copy of the Evaluation of the Honeywell Remote Oculometer Mark II Study has been provided to Secretary of the IRB; and Mr. Greg LaRosa, a member of the IRB representing the OCC, provided a copy of the wording for the informed consent form, which is determined to be suitable.

Based on the above resolution of the actions, the IRB determined that there are no problems associated with proceeding with the tests. If you have any questions, please feel free to call either the undersigned, or William M. Piland, Chairman of the IRB, at extension 44111.

C. Cockrell

Charles E. Cockrell
43361

cc:
106/Office of Director
106/W. M. Piland
421/OSEMA
141/OCC
141/G. C. LaRosa
152/CVIB
255A/R. A. Rivers
421/C. E. Cockrell

421/CECockrell:eat 8-20-96 (43361)

HUMAN ENGINEERING METHODS EXPERIMENTAL SUBJECT VOLUNTARY CONSENT FORM

I understand the purpose of the research and the techniques to be used as explained by the investigators. I understand that the following measurement systems will be in use during the experimental session: (1) Eye-movement monitoring through a head band mounted camera system, and (2) EEG through an electrode cap worn on the head, (3) Peripheral skin temperature measured from a sensor attached to the back of the finger, and (4) Videotapes of the Eye-movement monitoring system scene camera and of the processed EEG and temperature strip chart displays.

I am aware that the experiment eye-tracker utilizes an infrared illuminator (LED type, maximum energy of 0.8 milliwatts/cm^2 at the plane of the eye) which when directed at the retina may produce heat. I have been informed that the infrared exposure levels involved are well within known safety limits. No studies exist which have shown that any harm results from these levels of exposure.

I also understand that I am assured anonymity when the results are summarized and at any time I may withdraw from the experiment without further consequences to me. I understand that there are no known or expected physical or psychological side effects of this research. I do voluntarily consent to participate as a subject in the experiment as it is described to me.

________________________ ____________

(signature) *(date)*

____

(please print name)

____

Investigator

Point of Contact at NASA Langley:
J. Raymond Comstock, Jr. (757-864-6643)
Crew / Vehicle Integration Branch
NASA Langley Research Center

INFORMATION ABOUT THE PEACHES EXPERIMENT

The purpose of this research is to observe psychophysiological and eye-movement signs of alertness and attention as a function of performing a flight simulation task for an extended period of time.

Prior to the experimental session, a sensor cap will be placed on the subject's head to permit recording of brain wave, electroencephalogram (EEG) activity. The cap consists of 12 recessed sensors arranged in a standard placement system. It will be held in place by a chin strap. Adhesive sponge pads will be attached to the inside of the cap for comfort. Once the cap is in place, a dispenser tube with a hollow, blunt tip will be used to fill each of the sensors with conductive gel. Some slight abrading of the scalp with the blunted tip will be necessary to improve the sensor contact. Sensors held with adhesive pads and filled with conductive gel will be placed on the earlobes as reference points for the scalp sensors. There will be minimal discomfort associated with the sensor placement technique. The standard method of placement will include some slight abrasion or roughing of the skin at each location. There are no known side effects related to placement, except for slight redness which may occur subsequently, depending on the sensitivity of the skin.

Following sensor placement, the subject will then be fitted with a headband-mounted oculometer system to permit recording of eye movements. The headband may require repositioning to maintain an adequate comfort level. Please inform the experimenter if you experience any discomfort. A beam of near-infrared light will be directed at the left eye and a small video camera will pick up reflected light from this source. The subject will be asked to look at different areas on the console as the eye look-points are calibrated in the computer. To facilitate use of this eye tracking technique, the light emitting diode (LED) infrared illuminator will be directed toward your eye from the headband mounted oculometer system. The infrared energy seen by your eye is significantly less than your eye would see as ambient background light on a normal sunny day. This eye tracking technique has been used on hundreds of other subjects, and there are no documented or alleged cases of injury resulting from this eye tracking technique.

In addition to the above, peripheral skin temperature measurements will be made from a small sensor taped to the back of the left middle finger. Videotapes will be made of the Eye-movement monitoring system scene camera and of the processed EEG and temperature strip chart displays. Data from the experiment will be identified only by an assigned number, and not by your name to insure confidentiality.

Excluding the initial arrival briefing, the experimental sessions will last about seven hours (including a lunch break). The subject can, at any time without penalty, discontinue participation in the study.

Questions regarding the conduct of this experiment may be directed to J. R. Comstock, Jr. (Crew-Vehicle Integration Branch, MS-152, NASA LaRC, phone: 757-864-6643). Subjects concerned about protocol violations may request a meeting with the relevant NASA Langley Institutional Review Board (IRB).

APPENDIX B

PRACTICE COMPUTATIONAL SHEETS

Arrival Time

Seg	TOF(Hrs+min)	ETA
Takeoff	0 + 00	
1	0 + 30	
2	0 + 25	
3	0 + 20	
4	0 + 10	
5	0 + 10	

Time of Flight

Seg	TAS (kts)	Leg Dist(NM)	Time
1	180	90	
2	210	70	
3	240	100	
4	180	30	
5	150	25	

TOTAL ______________

Hrs + Min

Alternate Airfield Information

Heading - 260°

Distance - 120 NM

Divert Airspeed - 240 KTAS

Fuel Flow - 18000 lb/hour

Initial turn from current heading, direct to alternate will be left / right ? ________

The turn will be how many degrees?____

Time to fly to alternate will be? _______

Fuel used to alternate will be? _____ lbs

Landing Airspeed

Takeoff time - ________

Time Airborne - _______

Fuel Flow - 400 lbs/min

Takeoff Weight - 450,000 lbs

Fuel Consumed - _______ lbs

Current Weight - _______ lbs

Weight Range(lbs)	Airspeed
450,000-440,001	142 Kts
440,000-430,001	141 Kts
430,000-420,001	140 Kts
420,000-410,001	139 Kts

Weight and Balance	Subtotal (lbs)
Basic Aircraft Weight	250,000
Gallons Fuel x lbs/gallon 12,000 x 5	
# of Pax x Passenger Wt 100 x 160	
# of Pax x Baggage Wt 100 x 60	
Commercial cargo weight	68,000
Total (lbs)	

APPENDIX C

SIMULATION SESSION ONE SCRIPT

Performance Effects of Awareness Characterized by Hazardous and Effective States (PEACHES) - Phase I Simulator profile narrative and event outline

PEACHES will employ an instrument profile pattern for Runway 35L at the Denver International Airport. PAR, ILS and TACAN Approaches must be available. The instrument pattern flown will be a box pattern with a 16 NM downwind, and an 8 NM base leg. Pattern airspeed will be approximately 250 KTAS, and final airspeed is assumed to be approximately 180 KTAS. If a lower final airspeed is normal for the ACTS simulator, pattern airspeed and distances should be reduced to maintain a 12 minute pattern. The ACTS must also have an operational autopilot, and the ability to perform an autopilot coupled approach to 200' AGL.

The box pattern will be divided into two portions. The first includes a straight ahead climbout (4 NM), crosswind (8 NM), and downwind (16 NM). The second portion includes base (8 NM), final (10 NM), and missed approach to the end of the runway (2 NM). Each portion of the instrument pattern is designed to be six minutes in length.

Event times are listed below with the appropriate events for Simulator 1. Although Simulator 2 was originally planned to be a repeat of Sim 1, it was changed to prevent a confounding of study conditions. In the morning simulator the base/approach events will be manipulated to obtain the appropriate level (L - Low, B - Baseline, H- High) for the Index of Engagement (IE), while the crosswind/downwind events will attempt to maintain a baseline (medium) level IE. In the afternoon simulator, the crosswind/downwind events will be manipulated to obtain the targeted High and Low IEs, while the base/approach portion will serve as a baseline.

I. Simulator Period 1 - Profile Events	Time	IE
1. Startup - Performed by right seat	0+00 - 0+06	L
1. Subject monitors engine instruments	:00 - :06	
2. Subject performs Weight and Balance Exercise	:02 - :04	
3. Subject requests taxi/changes radio	:04	
(tone at mic key + every minute for entire study)		
2. Taxi - Performed by subject (External visual reference required)	0+06 - 0+12	B
1. Subject monitors engine instruments/checks navaids	:06 - :12	
2. Subject performs Time of Flight Exercise	:08 - :10	
3. Subject requests T/O clearanace/changes radio	:10	
3. Takeoff\Climbout - Performed by subject (All visuals lost)	0+12 - 0+18	H
1. Instrument T/O - Visual lost at 100'	:12 - :18	
2. Subject performes Arrival Time Exercise	:14 - :16	
3. Subject calls clearing 500'/changes radio	:16	
4. No Vectors, Std Inst Departure through turn to X-wind		
4. Pattern 1 crosswind/downwind	0+18 - 0+24	B
1. Instrument pattern 3000' AGL flown on autopilot	:18 - :24	
2. Subject performs Alternate Airfield Exercise	:20 - :22	
3. Subject changes radio/requests approaches	:22	
4. No vectors		

	Time	IE
5. Pattern 1 base/final - Demo by right seat	0+24 - 0+30	L
1. Descent to 1500' followed by ILS to 200'	:24 - :30	
2. Subject performs Landing Airspeed Exercise	:24 - :26	
3. Subject confirms gear down/cleared approach/changes radio	:28	
4. No vectors		
6. Missed Approach/Pattern 2 crosswind/downwind	0+30 - 0+36	B
1. Hand Flown - Controller Directed instrument pattern	:30 - :36	
2. Subject performs Alternate Airfield Exercise	:32 - :34	
3. Subject changes radio/ requests approach	:34	
4. No vectors		
7. Pattern 2 base/final - LOC Apprch with high, then low xwind/turb	0+36 - 0+42	H
1. Hand flown with controller inputs overcorrecting course	:36 - :42	
2. Subject performs Landing Airspeed Exercise	:36 - :38	
3. Subject confirms gear down/cleared approach/changes radio	:40	
8. Missed Approach/Pattern 3 crosswind/downwind	0+42 - 0+48	B
1. Right seat flown using autopilot	:42 - :48	
2. Subject performs Alternate Airfield Exercise	:44 - :46	
3. Subject changes radio/requests approach	:46	

9. Pattern 3 base/final - ILS no wind	0+48 - 0+54	L
1. Right seat flown using autopilot	:48 - :54	
2. Subject performs Landing Airspeed Exercise	:48 - :50	
3. Subject confirms gear down/cleared approach/changes radio	:52	
4. Vectors after base turn until after final		
10. Missed Approach/Pattern 4 crosswind/downwind	0+54 - 1+00	B
1. Hand Flown - Controller Directed instrument pattern	:54 - 1:00	
2. Subject performs Alternate Airfield Exercise	:56 - :58	
3. Subject changes radio/ requests approach/ sets altimeter	:58	
4. Vectors after turn to crosswind		
11. Pattern 4 base/final - PAR with xwinds and turbulence	1+00 - 1+06	H
1. Hand flown by subject	1:00 - 1:06	
2. Subject performs Landing Airspeed Exercise	1:00 - 1:02	
3. Subject confirms gear down/cleared approach/changes radio	1:04	
4. No vectors after crossing threshold		
12. Missed Approach/Pattern 5 crosswind/downwind	1+06 - 1+12	B
1. Hand Flown - Controller Directed instrument pattern	1:06 - 1:12	
2. Subject performs Alternate Airfield Exercise	1:08 - 1:10	
3. Subject changes radio/ requests & sets up approach	1:10	
4. No vectors		

17. Pattern 7 base/final - ILS to 200'	1+36 - 1+42 L
1. Right seat hand flown, subject monitored	1:36 - 1:42
2. Subject performs Landing Airspeed Exercise	1:36 - 1:38
3. Subject confirms gear down/cleared approach/changes radio	1:40
4. No vectors	
18. Missed Approach/Pattern 8 crosswind/downwind	1+42 - 1+48 B
1. Subject Hand Flown - Controller Directed instrument pattern	1:42 - 1:48
2. Subject performs Alternate Airfield Exercise	1:44 - 1:46
3. Subject changes radio/ requests approach	1:46
4. Vectors after turn to crosswind	
19. Pattern 8 base/final - ILS to 200' with rapid fuel depletion (IFE)	1+48 - 1+54 H
1. Subject flown/monitored	1:48 - 1:54
2. Subject performs Landing Airspeed Exercise	1:48 - 1:50
3. Subject confirms gear down/cleared approach/changes radio	1:52
4. No vectors	

Break for lunch, second simulator period begins with simulator airborne in this position.

13. Pattern 5 base/final - ILS/autopilot coupled approach to 200' 1+12 - 1+18 L

    1. Subject flown/monitored 1:12 - 1:18

    2. Subject performs Landing Airspeed Exercise 1:12 - 1:14

    3. Subject confirms gear down/cleared approach/changes radio 1:16

    4. Vectors after turn to base until final

14. Missed Approach/Pattern 6 crosswind/downwind 1+18 - 1+24 B

    1. Subject Hand Flown - Controller Directed instrument pattern 1:18 - 1:24

    2. Subject performs Alternate Airfield Exercise 1:20 - 1:22

    3. Subject changes radio/ requests approach 1:18

    4. Vectors after turn to crosswind

15. Pattern 6 base/final - PAR with speed changes for traffic/ + xwind 1+24 - 1+30 H

    1. Subject Hand Flown, directed by overzealous controller 1:24 - 1:30

    2. Subject performs Landing Airspeed Exercise 1:24 - 1:26

    3. Subject confirms gear down/cleared approach/changes radio 1:28

    4. No vectors after crossing the threshold

16. Missed Approach/Pattern 7 crosswind/downwind 1+30 - 1+36 B

    1. Subject flown using autopilot 1:30 - 1.36

    2. Subject performs Alternate Airfield Exercise 1:32 - 1:34

    3. Subject changes radio/requests approach 1:34

    4. Vectors after turn to crosswind

APPENDIX D

SIMULATOR SESSION TWO SCRIPT

II. Simulator Period 2 - Profile Events

	Time	IE
1. Missed Approach/Pattern 9 crosswind/downwind		
1. Right seat flown using autopilot	See Sim Period 1	L
2. Subject performs Alternate Airfield Exercise		
3. Subject changes radio/requests approach		
4. No vectors		
2. Pattern 9 base/final - ILS to 200'		
1. Subject flown/monitored		B
2. Subject performs Landing Airspeed Exercise		
3. Subject confirms gear down/cleared approach/changes radio		
3. Missed Approach/Pattern 10 crosswind/downwind		
1. Controller directed pattern with high wind (no turbulence) and **multiple heading changes**		H
2. Subject performs Alternate Airfield Exercise		
3. Subject changes radio/requests approach		
4. Pattern 10 base/final - ILS to 200'		
1. Subject flown/monitored		B
2. Subject performs Landing Airspeed Exercise		
3. Subject confirms gear down/cleared approach/changes radio		
4. No vectors on missed approacn		

5. Missed Approach/Pattern 11 crosswind/downwind L
    1. Right seat flown using autopilot
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach
    4. No vectors
6. Pattern 11 base/final - ILS to 200' B
    1. Subject flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio
7. Missed Approach/Pattern 12 crosswind/downwind H
    1. Controller directed patterns with high wind (no turbulence) and multiple altitude changes
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach
8. Pattern 12 base/final - ILS to 200' B
    1. Subject flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio
    4. No vectors on missed approach

9. Missed Approach/Pattern 13 crosswind/downwind L
    1. Autopilot coupled subject monitored
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach
    4. No vectors
10. Pattern 13 base/final - ILS to 200' B
    1. Right seat flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio
    4. No vectors
11. Missed Approach/Pattern 14 crosswind/downwind L
    1. Autopilot coupled right seat monitored
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach
    4. No vectors
12. Pattern 14 base/final - ILS to 200' B
    1. Subject flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio

13. Missed Approach/Pattern 15 crosswind/downwind H

    1. Subject flown, controller directed with **multiple heading changes**

    2. Subject performs Alternate Airfield Exercise

    3. Subject changes radio/requests approach

14. Pattern 15 base/final - ILS to 200' B

    1. Subject flown/monitored

    2. Subject performs Landing Airspeed Exercise

    3. Subject confirms gear down/cleared approach/changes radio

15. Missed Approach/Pattern 16 crosswind/downwind H

    1. Subject flown, controller directed with **multiple altitude changes**

    2. Subject performs Alternate Airfield Exercise

    3. Subject changes radio/requests approach

16. Pattern 16 base/final - ILS to 200' B

    1. Subject flown/monitored

    2. Subject performs Landing Airspeed Exercise

    3. Subject confirms gear down/cleared approach/changes radio

    4. No vectors on missed approach

17. Missed Approach/Pattern 17 crosswind/downwind L

    1. Autopilot coupled right seat monitored
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach
    4. No vectors

18. Pattern 17 base/final - ILS to 200' B

    1. Subject flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio

19. Missed Approach/Pattern 18 crosswind/downwind H

    1. Subject flown, controller directed with **multiple heading and altitude changes**
    2. Subject performs Alternate Airfield Exercise
    3. Subject changes radio/requests approach

20. Pattern 18 base/final - ILS to landing B

    1. Subject flown/monitored
    2. Subject performs Landing Airspeed Exercise
    3. Subject confirms gear down/cleared approach/changes radio

APPENDIX E

ANOVA TABLES

Table E1. Five Factor ANOVA for Airspeed Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	219.34	219.34	T x S(G)	2.23	NS
Group	[G]	3	249.16	83.05	S(G)	0.49	NS
#Subject(Group)	[S(G)]	12	8098.22	674.85		4.02***	
Time Day	[D]	1	0.13	0.13	D x S(G)	0.00	NS
Workload Level	[W]	2	4484.86	2242.43	W x S(G)	33.75***	0.202
Interactions							
T x G		3	292.49	97.50	T x S(G)	0.99	NS
T x D		1	5.92	5.92	T x D x S(G)	0.29	NS
T x W		2	79.67	39.83	T x W x S(G)	0.97	NS
G x D		3	135.72	45.24	D x S(G)	0.43	NS
G x W		6	246.52	41.09	W x S(G)	0.62	NS
D x W		2	3013.12	1506.56	D x W x S(G)	18.09***	0.132
T x G x D		3	155.67	51.89	T x D x S(G)	2.52	NS
T x G x W		6	380.86	63.48	T x W x S(G)	1.55	NS
T x D x W		2	768.46	384.23	T x D x W x S(G)	8.45**	0.032
G x D x W		6	707.23	117.87	D x W x S(G)	1.42	NS
T x G x D x W		6	334.75	55.79	T x D x W x S(G)	1.23	NS
Error Terms							
S(G)		12	2024.56	168.71			
T x S(G)		12	1179.98	98.33			
D x S(G)		12	1258.12	104.84			
W x S(G)		24	1594.45	66.44			
T x D x S(G)		12	247.27	20.61			
T x W x S(G)		24	985.79	41.07			
D x W x S(G)		24	1998.58	83.27			
T x D x W x S(G)		24	1091.40	45.48			
Total		191	21454.05				

* = p<0.05   ** = p<0.01   *** = p<0.001   NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E2. Five Factor ANOVA for Composite Airspeed Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	11.72	11.72	T x S(G)	3.29	NS
Group	[G]	3	13.92	4.64	S(G)	0.94	NS
#Subject(Group)	[S(G)]	12	236.67	19.72		3.83***	
Time Day	[D]	1	1.27	1.27	D x S(G)	0.41	NS
Workload Level	[W]	2	121.80	60.90	W x S(G)	25.76***	0.140
Interactions							
T x G		3	11.11	3.70	T x S(G)	1.04	NS
T x D		1	0.18	0.18	T x D x S(G)	0.23	NS
T x W		2	4.16	2.08	T x W x S(G)	1.68	NS
G x D		3	7.07	2.36	D x S(G)	0.75	NS
G x W		6	6.88	1.15	W x S(G)	0.49	NS
D x W		2	228.40	114.20	D x W x S(G)	38.62***	0.265
T x G x D		3	4.16	1.39	T x D x S(G)	1.76	NS
T x G x W		6	11.54	1.92	T x W x S(G)	1.55	NS
T x D x W		2	25.20	12.60	T x D x W x S(G)	6.51**	0.025
G x D x W		6	20.69	3.45	D x W x S(G)	1.17	NS
T x G x D x W		6	14.83	2.47	T x D x W x S(G)	1.28	NS
Error Terms							
S(G)		12	59.17	4.93			
T x S(G)		12	42.77	3.56			
D x S(G)		12	37.68	3.14			
W x S(G)		24	56.73	2.36			
T x D x S(G)		12	9.47	0.79			
T x W x S(G)		24	29.72	1.24			
D x W x S(G)		24	70.97	2.96			
T x D x W x S(G)		24	46.48	1.94			
Total		191	835.91				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E3. Five Factor ANOVA for Composite Cross Track Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	148.55	148.55	T x S(G)	71.16***	0.089
Group	[G]	3	4.57	1.52	S(G)	0.35	NS
#Subject(Group)	[S(G)]	12	206.22	17.19		1.26	
Time Day	[D]	1	74.49	74.49	D x S(G)	19.76***	0.043
Workload Level	[W]	2	185.88	92.94	W x S(G)	25.46***	0.108
Interactions							
T x G		3	11.62	3.87	T x S(G)	1.86	NS
T x D		1	41.62	41.62	T x D x S(G)	15.60**	0.024
T x W		2	116.89	58.45	T x W x S(G)	23.00***	0.068
G x D		3	2.51	0.84	D x S(G)	0.22	NS
G x W		6	2.90	0.48	W x S(G)	0.13	NS
D x W		2	293.50	146.75	D x W x S(G)	40.78***	0.174
T x G x D		3	9.24	3.08	T x D x S(G)	1.15	NS
T x G x W		6	21.41	3.57	T x W x S(G)	1.40	NS
T x D x W		2	237.56	118.78	T x D x W x S(G)	45.23***	0.141
G x D x W		6	19.08	3.18	D x W x S(G)	0.88	NS
T x G x D x W		6	21.11	3.52	T x D x W x S(G)	1.34	NS
Error Terms							
S(G)		12	51.56	4.30			
T x S(G)		12	25.05	2.09			
D x S(G)		12	45.23	3.77			
W x S(G)		24	87.60	3.65			
T x D x S(G)		12	32.01	2.67			
T x W x S(G)		24	60.99	2.54			
D x W x S(G)		24	86.37	3.60			
T x D x W x S(G)		24	63.03	2.63			
Total		191	1642.78				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E4. Five Factor ANOVA for Composite Vertical Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	18.85	18.85	T x S(G)	11.65**	0.013
Group	[G]	3	11.30	3.77	S(G)	1.09	NS
#Subject(Group)	[S(G)]	12	166.16	13.85		1.02	NS
Time Day	[D]	1	8.97	8.97	D x S(G)	2.70	NS
Workload Level	[W]	2	34.81	17.40	W x S(G)	9.24**	0.023
Interactions							
T x G		3	20.23	6.74	T x S(G)	4.17*	0.012
T x D		1	69.15	69.15	T x D x S(G)	34.34***	0.051
T x W		2	122.49	61.24	T x W x S(G)	26.60***	0.089
G x D		3	1.53	0.51	D x S(G)	0.15	NS
G x W		6	7.45	1.24	W x S(G)	0.66	NS
D x W		2	535.76	267.88	D x W x S(G)	112.72***	0.401
T x G x D		3	3.84	1.28	T x D x S(G)	0.64	NS
T x G x W		6	14.64	2.44	T x W x S(G)	1.06	NS
T x D x W		2	97.56	48.78	T x D x W x S(G)	24.32***	0.071
G x D x W		6	23.11	3.85	D x W x S(G)	1.62	NS
T x G x D x W		6	20.67	3.44	T x D x W x S(G)	1.72	NS
Error Terms							
S(G)		12	41.54	3.46			
T x S(G)		12	19.42	1.62			
D x S(G)		12	39.83	3.32			
W x S(G)		24	45.20	1.88			
T x D x S(G)		12	24.16	2.01			
T x W x S(G)		24	55.25	2.30			
D x W x S(G)		24	57.04	2.38			
T x D x W x S(G)		24	48.13	2.01			
Total		191	1320.94				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E5. Five Factor ANOVA for Adjusted Performance Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	34.92	34.92	T x S(G)	6.85*	0.023
Group	[G]	3	6.64	2.21	S(G)	0.27	NS
#Subject(Group)	[S(G)]	12	394.06	32.84		4.31***	
Time Day	[D]	1	3.36	3.36	D x S(G)	0.62	NS
Workload Level	[W]	2	472.12	236.06	W x S(G)	71.63***	0.358
Interactions							
T x G		3	16.68	5.56	T x S(G)	1.09	NS
T x D		1	1.25	1.25	T x D x S(G)	0.80	NS
T x W		2	17.11	8.56	T x W x S(G)	4.44*	0.010
G x D		3	2.48	0.83	D x S(G)	0.15	NS
G x W		6	13.59	2.27	W x S(G)	0.69	NS
D x W		2	39.86	19.93	D x W x S(G)	4.47*	0.024
T x G x D		3	5.27	1.76	T x D x S(G)	1.12	NS
T x G x W		6	23.44	3.91	T x W x S(G)	2.03	NS
T x D x W		2	50.29	25.15	T x D x W x S(G)	9.56***	0.035
G x D x W		6	52.52	8.75	D x W x S(G)	1.97	NS
T x G x D x W		6	17.48	2.91	T x D x W x S(G)	1.11	NS
Error Terms							
S(G)		12	98.52	8.21			
T x S(G)		12	61.16	5.10			
D x S(G)		12	64.54	5.38			
W x S(G)		24	79.10	3.30			
T x D x S(G)		12	18.79	1.57			
T x W x S(G)		24	46.28	1.93			
D x W x S(G)		24	106.91	4.45			
T x D x W x S(G)		24	63.16	2.63			
Total		191	1295.48				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E6. Five Factor ANOVA for Composite Error

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	365.51	365.51	T x S(G)	30.78***	0.048
Group	[G]	3	34.76	11.59	S(G)	0.58	NS
#Subject(Group)	[S(G)]	12	967.08	80.59		2.14*	
Time Day	[D]	1	6.53	6.53	D x S(G)	0.33	NS
Workload Level	[W]	2	1526.11	763.06	W x S(G)	50.27***	0.203
Interactions							
T x G		3	112.63	37.54	T x S(G)	3.16	NS
T x D		1	172.03	172.03	T x D x S(G)	26.79***	0.023
T x W		2	544.28	272.14	T x W x S(G)	31.96***	0.072
G x D		3	2.28	0.76	D x S(G)	0.04	NS
G x W		6	7.00	1.17	W x S(G)	0.08	NS
D x W		2	1748.15	874.08	D x W x S(G)	64.40***	0.234
T x G x D		3	12.94	4.31	T x D x S(G)	0.67	NS
T x G x W		6	90.43	15.07	T x W x S(G)	1.77	NS
T x D x W		2	676.57	338.28	T x D x W x S(G)	32.62***	0.089
G x D x W		6	77.93	12.99	D x W x S(G)	0.96	NS
T x G x D x W		6	137.46	22.91	T x D x W x S(G)	2.21	NS
Error Terms							
S(G)		12	241.77	20.15			
T x S(G)		12	142.52	11.88			
D x S(G)		12	234.50	19.54			
W x S(G)		24	364.27	15.18			
T x D x S(G)		12	77.07	6.42			
T x W x S(G)		24	204.38	8.52			
D x W x S(G)		24	325.72	13.57			
T x D x W x S(G)		24	248.90	10.37			
Total		191	7353.74				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E7. Five Factor ANOVA for Performance Error Rating

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.4701	0.4701	T x S(G)	2.67	NS
Group	[G]	3	0.2956	3.3385	S(G)	0.35	NS
#Subject(Group)	[S(G)]	12	13.3542	209.5000		3.06	
Time Day	[D]	1	13.2826	1.9427	D x S(G)	82.05	0.144
Workload Level	[W]	2	24.5000	2.0755	W x S(G)	141.65***	0.267
Interactions							
T x G		3	1.6706	2.1094	T x S(G)	3.17	NS
T x D		1	1.4180	1.0365	T x D x S(G)	16.42***	0.015
T x W		2	6.4245	2.6953	T x W x S(G)	28.60***	0.068
G x D		3	0.6706	1.9427	D x S(G)	1.38	NS
G x W		6	0.6536	2.0755	W x S(G)	1.26	NS
D x W		2	18.0339	3.4870	D x W x S(G)	62.06***	0.195
T x G x D		3	0.0456	1.0365	T x D x S(G)	0.18	NS
T x G x W		6	0.9427	2.6953	T x W x S(G)	1.40	NS
T x D x W		2	3.2578	1.0339	T x D x W x S(G)	37.81***	0.035
G x D x W		6	0.3958	3.4870	D x W x S(G)	0.45	NS
T x G x D x W		6	1.2083	1.0339	T x D x W x S(G)	4.68	0.010
Error Terms							
S(G)		12	3.3385	0.2782			
T x S(G)		12	2.1094	0.1758			
D x S(G)		12	1.9427	0.1619			
W x S(G)		24	2.0755	0.0864			
T x D x S(G)		12	1.0365	0.0864			
T x W x S(G)		24	2.6953	0.1123			
D x W x S(G)		24	3.4870	0.1453			
T x D x W x S(G)		24	1.0339	0.0430			
Total		191	90.9883				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E8. Five Factor ANOVA for Pupil Diameter

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	3968.26	3968.26	T x S(G)	19.73***	0.028
Group	[G]	3	16437.61	5479.20	S(G)	0.78	NS
#Subject(Group)	[S(G)]	12	339232.48	28269.37		303.02***	
Time Day	[D]	1	666.58	666.58	D x S(G)	0.90	NS
Workload Level	[W]	3	1273.00	424.33	W x S(G)	8.90***	0.009
Interactions							
T x G		3	1994.62	664.87	T x S(G)	3.31	NS
T x D		1	49.89	49.89	T x D x S(G)	0.26	NS
T x W		3	463.41	154.47	T x W x S(G)	6.89***	0.003
G x D		3	898.58	299.53	D x S(G)	0.41	NS
G x W		9	562.01	62.45	W x S(G)	1.31	NS
D x W		3	5.51	1.84	D x W x S(G)	0.05	NS
T x G x D		3	760.42	253.47	T x D x S(G)	1.33	NS
T x G x W		9	520.66	57.85	T x W x S(G)	2.58*	0.002
T x D x W		3	291.09	97.03	T x D x W x S(G)	2.78	NS
G x D x W		9	692.74	76.97	D x W x S(G)	1.90	NS
T x G x D x W		9	457.84	50.87	T x D x W x S(G)	1.46	NS
Error Terms							
S(G)		12	84808.12	7067.34			
T x S(G)		12	2412.95	201.08			
D x S(G)		12	8873.21	739.43			
W x S(G)		36	1716.25	47.67			
T x D x S(G)		12	2285.07	190.42			
T x W x S(G)		36	807.07	22.42			
D x W x S(G)		36	1460.79	40.58			
T x D x W x S(G)		36	1257.72	34.94			
Total		255	132663.40				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E9. Five Factor ANOVA for Pupil Diameter Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	30.92	30.92	T x S(G)	3.04	NS
Group	[G]	3	41.78	13.93	S(G)	1.40	NS
#Subject(Group)	[S(G)]	12	478.76	39.90		0.42	
Time Day	[D]	1	6.51	6.51	D x S(G)	0.48	NS
Workload Level	[W]	3	2205.80	735.27	W x S(G)	9.68***	0.128
Interactions							
T x G		3	33.96	11.32	T x S(G)	1.11	NS
T x D		1	2.33	2.33	T x D x S(G)	0.17	NS
T x W		3	582.15	194.05	T x W x S(G)	6.03**	0.031
G x D		3	47.92	15.97	D x S(G)	1.18	NS
G x W		9	561.87	62.43	W x S(G)	0.82	NS
D x W		3	19.64	6.55	D x W x S(G)	0.12	NS
T x G x D		3	56.80	18.93	T x D x S(G)	1.40	NS
T x G x W		9	736.99	81.89	T x W x S(G)	2.54*	0.029
T x D x W		3	567.49	189.16	T x D x W x S(G)	3.06*	0.025
G x D x W		9	1097.31	121.92	D x W x S(G)	2.19*	0.039
T x G x D x W		9	718.76	79.86	T x D x W x S(G)	1.29	NS
Error Terms							
S(G)		12	119.69	9.97			
T x S(G)		12	122.02	10.17			
D x S(G)		12	162.89	13.57			
W x S(G)		36	2734.64	75.96			
T x D x S(G)		12	161.76	13.48			
T x W x S(G)		36	1159.16	32.20			
D x W x S(G)		36	2006.57	55.74			
T x D x W x S(G)		36	2228.38	61.90			
Total		255	15405.34				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E10. Five Factor ANOVA for Peripheral Temperature

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.021	0.021	T x S(G)	0.04	NS
Group	[G]	3	1133.192	377.731	S(G)	1.93	NS
#Subject(Group)	[S(G)]	12	9381.728	781.811		82.41***	
Time Day	[D]	1	546.420	546.420	D x S(G)	12.30**	0.089
Workload Level	[W]	3	477.028	159.009	W x S(G)	55.03***	0.084
Interactions							
T x G		3	0.677	0.226	T x S(G)	0.46	NS
T x D		1	0.057	0.057	T x D x S(G)	0.35	NS
T x W		3	12.719	4.240	T x W x S(G)	11.15***	0.002
G x D		3	140.386	46.795	D x S(G)	1.05	NS
G x W		9	78.360	8.707	W x S(G)	3.01**	0.009
D x W		3	113.040	37.680	D x W x S(G)	40.35***	0.020
T x G x D		3	0.602	0.201	T x D x S(G)	1.24	NS
T x G x W		9	6.142	0.682	T x W x S(G)	1.79	NS
T x D x W		3	5.461	1.820	T x D x W x S(G)	6.54**	0.001
G x D x W		9	19.821	2.202	D x W x S(G)	2.36*	0.002
T x G x D x W		9	2.803	0.311	T x D x W x S(G)	1.12	NS
Error Terms							
S(G)		12	2345.432	195.453			
T x S(G)		12	5.872	0.489			
D x S(G)		12	533.041	44.420			
W x S(G)		36	104.013	2.889			
T x D x S(G)		12	1.951	0.163			
T x W x S(G)		36	13.689	0.380			
D x W x S(G)		36	33.619	0.934			
T x D x W x S(G)		36	10.019	0.278			
Total		255	5584.366				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E11. Five Factor ANOVA for Peripheral Temp Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.0885	0.0885	T x S(G)	2.07	NS
Group	[G]	3	0.3061	0.1020	S(G)	0.14	NS
#Subject(Group)	[S(G)]	12	34.1114	2.8426		0.55	
Time Day	[D]	1	2.6596	2.6596	D x S(G)	6.30*	0.002
Workload Level	[W]	3	613.4975	204.4992	W x S(G)	38.54***	0.464
Interactions							
T x G		3	0.0084	0.0028	T x S(G)	0.07	NS
T x D		1	0.2048	0.2048	T x D x S(G)	5.29*	0.000
T x W		3	20.4355	6.8118	T x W x S(G)	7.09***	0.014
G x D		3	1.1551	0.3850	D x S(G)	0.91	NS
G x W		9	112.0850	12.4539	W x S(G)	2.35*	0.050
D x W		3	119.2172	39.7391	D x W x S(G)	24.90***	0.089
T x G x D		3	0.0953	0.0318	T x D x S(G)	0.82	NS
T x G x W		9	10.7726	1.1970	T x W x S(G)	1.24	NS
T x D x W		3	18.5788	6.1929	T x D x W x S(G)	8.77***	0.013
G x D x W		9	53.0264	5.8918	D x W x S(G)	3.69**	0.030
T x G x D x W		9	6.9111	0.7679	T x D x W x S(G)	1.09	NS
Error Terms							
S(G)		12	8.5278	0.7107			
T x S(G)		12	0.5131	0.0428			
D x S(G)		12	5.0683	0.4224			
W x S(G)		36	190.9999	5.3056			
T x D x S(G)		12	0.4644	0.0387			
T x W x S(G)		36	34.6117	0.9614			
D x W x S(G)		36	57.4640	1.5962			
T x D x W x S(G)		36	25.4329	0.7065			
Total		255	1282.1241				

* = $p<0.05$ ** = $p<0.01$ *** = $p<0.001$ NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E12. Five Factor ANOVA for Saccade Time

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.000593	0.000593	T x S(G)	42.35***	0.128
Group	[G]	3	0.000426	0.000142	S(G)	0.94	NS
#Subject(Group)	[S(G)]	12	0.007264	0.000605		68.13***	
Time Day	[D]	1	0.000020	0.000020	D x S(G)	0.63	NS
Workload Level	[W]	3	0.000156	0.000052	W x S(G)	13.23***	0.032
Interactions							
T x G		3	0.000068	0.000023	T x S(G)	1.63	NS
T x D		1	0.000000	0.000000	T x D x S(G)	0.00	NS
T x W		3	0.000016	0.000005	T x W x S(G)	1.93	NS
G x D		3	0.000061	0.000020	D x S(G)	0.63	NS
G x W		9	0.000028	0.000003	W x S(G)	0.80	NS
D x W		3	0.000068	0.000023	D x W x S(G)	7.47***	0.013
T x G x D		3	0.000042	0.000014	T x D x S(G)	2.08	NS
T x G x W		9	0.000042	0.000005	T x W x S(G)	1.66	NS
T x D x W		3	0.000019	0.000006	T x D x W x S(G)	2.21	NS
G x D x W		9	0.000026	0.000003	D x W x S(G)	0.97	NS
T x G x D x W		9	0.000019	0.000002	T x D x W x S(G)	0.76	NS
Error Terms							
S(G)		12	0.001816	0.000151			
T x S(G)		12	0.000168	0.000014			
D x S(G)		12	0.000389	0.000032			
W x S(G)		36	0.000142	0.000004			
T x D x S(G)		12	0.000080	0.000007			
T x W x S(G)		36	0.000101	0.000003			
D x W x S(G)		36	0.000109	0.000003			
T x D x W x S(G)		36	0.000102	0.000003			
Total		255	0.004493				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E13. Five Factor ANOVA for Saccade Time Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.0000008	0.0000008	T x S(G)	1.91	NS
Group	[G]	3	0.0000028	0.0000009	S(G)	2.18	NS
#Subject(Group)	[S(G)]	12	0.0000202	0.0000017		0.10	
Time Day	[D]	1	0.0000044	0.0000044	D x S(G)	8.20*	0.002
Workload Level	[W]	3	0.0002029	0.0000676	W x S(G)	6.60**	0.095
Interactions							
T x G		3	0.0000018	0.0000006	T x S(G)	1.37	NS
T x D		1	0.0000000	0.0000000	T x D x S(G)	0.01	NS
T x W		3	0.0000360	0.0000120	T x W x S(G)	1.54	NS
G x D		3	0.0000023	0.0000008	D x S(G)	1.40	NS
G x W		9	0.0000635	0.0000071	W x S(G)	0.69	NS
D x W		3	0.0000563	0.0000188	D x W x S(G)	2.45	NS
T x G x D		3	0.0000014	0.0000005	T x D x S(G)	1.06	NS
T x G x W		9	0.0000797	0.0000089	T x W x S(G)	1.14	NS
T x D x W		3	0.0000329	0.0000110	T x D x W x S(G)	1.48	NS
G x D x W		9	0.0000526	0.0000058	D x W x S(G)	0.76	NS
T x G x D x W		9	0.0000499	0.0000055	T x D x W x S(G)	0.75	NS
Error Terms							
S(G)		12	0.0000050	0.0000004			
T x S(G)		12	0.0000052	0.0000004			
D x S(G)		12	0.0000064	0.0000005			
W x S(G)		36	0.0003687	0.0000102			
T x D x S(G)		12	0.0000052	0.0000004			
T x W x S(G)		36	0.0002796	0.0000078			
D x W x S(G)		36	0.0002751	0.0000076			
T x D x W x S(G)		36	0.0002660	0.0000074			
Total		255	0.0017983				

* = $p<0.05$ ** = $p<0.01$ *** = $p<0.001$ NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E14. Five Factor ANOVA for Saccade Distance

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	1042	1042	T x S(G)	2.65	NS
Group	[G]	3	2997	999	S(G)	0.31	NS
#Subject(Group)	[S(G)]	12	153356	12780		18.07***	
Time Day	[D]	1	3526	3526	D x S(G)	2.55	NS
Workload Level	[W]	3	899	300	W x S(G)	1.10	NS
Interactions							
T x G		3	1141	380	T x S(G)	0.97	NS
T x D		1	1344	1344	T x D x S(G)	4.03	NS
T x W		3	114	38	T x W x S(G)	0.33	NS
G x D		3	1923	641	D x S(G)	0.46	NS
G x W		9	3335	371	W x S(G)	1.36	NS
D x W		3	1937	646	D x W x S(G)	2.03	NS
T x G x D		3	470	157	T x D x S(G)	0.47	NS
T x G x W		9	1074	119	T x W x S(G)	1.03	NS
T x D x W		3	489	163	T x D x W x S(G)	1.74	NS
G x D x W		9	4271	475	D x W x S(G)	1.49	NS
T x G x D x W		9	636	71	T x D x W x S(G)	0.75	NS
Error Terms							
S(G)		12	38339	3195			
T x S(G)		12	4713	393			
D x S(G)		12	16566	1380			
W x S(G)		36	9773	271			
T x D x S(G)		12	4002	333			
T x W x S(G)		36	4180	116			
D x W x S(G)		36	11476	319			
T x D x W x S(G)		36	3385	94			
Total		255	117631				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E15. Five Factor ANOVA for Saccade Distance Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	93.32	93.32	T x S(G)	1.74	NS
Group	[G]	3	336.69	112.23	S(G)	0.99	NS
#Subject(Group)	[S(G)]	12	5459.26	454.94		0.76	
Time Day	[D]	1	143.99	143.99	D x S(G)	1.32	NS
Workload Level	[W]	3	2395.36	798.45	W x S(G)	2.61	NS
Interactions							
T x G		3	205.87	68.62	T x S(G)	1.28	NS
T x D		1	71.68	71.68	T x D x S(G)	1.23	NS
T x W		3	248.35	82.78	T x W x S(G)	0.29	NS
G x D		3	384.87	128.29	D x S(G)	1.17	NS
G x W		9	3743.91	415.99	W x S(G)	1.36	NS
D x W		3	6627.20	2209.07	D x W x S(G)	7.51***	0.084
T x G x D		3	195.53	65.18	T x D x S(G)	1.12	NS
T x G x W		9	2052.06	228.01	T x W x S(G)	0.81	NS
T x D x W		3	1271.84	423.95	T x D x W x S(G)	1.86	NS
G x D x W		9	4968.72	552.08	D x W x S(G)	1.88	NS
T x G x D x W		9	1701.88	189.10	T x D x W x S(G)	0.83	NS
Error Terms							
S(G)		12	1364.81	113.73			
T x S(G)		12	645.07	53.76			
D x S(G)		12	1310.91	109.24			
W x S(G)		36	11006.45	305.73			
T x D x S(G)		12	700.00	58.33			
T x W x S(G)		36	10154.86	282.08			
D x W x S(G)		36	10584.18	294.01			
T x D x W x S(G)		36	8197.42	227.71			
Total		255	68405.01				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E16. Five Factor ANOVA for Fixation Size

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.0551	0.0551	T x S(G)	2.56	NS
Group	[G]	3	0.1312	0.0437	S(G)	0.49	NS
#Subject(Group)	[S(G)]	12	4.3121	0.3593		68.33***	
Time Day	[D]	1	0.0311	0.0311	D x S(G)	3.83	NS
Workload Level	[W]	3	0.0125	0.0042	W x S(G)	1.83	NS
Interactions							
T x G		3	0.0267	0.0089	T x S(G)	0.41	NS
T x D		1	0.0109	0.0109	T x D x S(G)	1.51	NS
T x W		3	0.0714	0.0238	T x W x S(G)	8.19***	0.025
G x D		3	0.0817	0.0272	D x S(G)	3.36	NS
G x W		9	0.0401	0.0045	W x S(G)	1.96	NS
D x W		3	0.0752	0.0251	D x W x S(G)	13.31***	0.027
T x G x D		3	0.0064	0.0021	T x D x S(G)	0.29	NS
T x G x W		9	0.0222	0.0025	T x W x S(G)	0.85	NS
T x D x W		3	0.0340	0.0113	T x D x W x S(G)	4.86**	0.011
G x D x W		9	0.0384	0.0043	D x W x S(G)	2.27*	0.008
T x G x D x W		9	0.0387	0.0043	T x D x W x S(G)	1.84	NS
Error Terms							
S(G)		12	1.0780	0.0898			
T x S(G)		12	0.2585	0.0215			
D x S(G)		12	0.0974	0.0081			
W x S(G)		36	0.0820	0.0023			
T x D x S(G)		12	0.0867	0.0072			
T x W x S(G)		36	0.1046	0.0029			
D x W x S(G)		36	0.0678	0.0019			
T x D x W x S(G)		36	0.0840	0.0023			
Total		255	2.5349				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E17. Five Factor ANOVA for Fixation Size Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.00217	0.00217	T x S(G)	1.04	NS
Group	[G]	3	0.00487	0.00162	S(G)	0.82	NS
#Subject(Group)	[S(G)]	12	0.09503	0.00792		0.70	
Time Day	[D]	1	0.00027	0.00027	D x S(G)	0.37	NS
Workload Level	[W]	3	0.06093	0.02031	W x S(G)	4.19*	0.028
Interactions							
T x G		3	0.00697	0.00232	T x S(G)	1.12	NS
T x D		1	0.00066	0.00066	T x D x S(G)	0.62	NS
T x W		3	0.09955	0.03318	T x W x S(G)	4.60**	0.047
G x D		3	0.00095	0.00032	D x S(G)	0.43	NS
G x W		9	0.07080	0.00787	W x S(G)	1.62	NS
D x W		3	0.24055	0.08018	D x W x S(G)	16.92***	0.136
T x G x D		3	0.00146	0.00049	T x D x S(G)	0.46	NS
T x G x W		9	0.07038	0.00782	T x W x S(G)	1.08	NS
T x D x W		3	0.04174	0.01391	T x D x W x S(G)	2.37	NS
G x D x W		9	0.06012	0.00668	D x W x S(G)	1.41	NS
T x G x D x W		9	0.11539	0.01282	T x D x W x S(G)	2.19*	0.038
Error Terms							
S(G)		12	0.02376	0.00198			
T x S(G)		12	0.02499	0.00208			
D x S(G)		12	0.00882	0.00074			
W x S(G)		36	0.17438	0.00484			
T x D x S(G)		12	0.01281	0.00107			
T x W x S(G)		36	0.25990	0.00722			
D x W x S(G)		36	0.17060	0.00474			
T x D x W x S(G)		36	0.21091	0.00586			
Total		255	1.66299				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E18. Five Factor ANOVA for Ellipticity

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.018	0.018	T x S(G)	0.30	NS
Group	[G]	3	1.242	0.414	S(G)	1.20	NS
#Subject(Group)	[S(G)]	12	16.554	1.379		72.32***	
Time Day	[D]	1	0.125	0.125	D x S(G)	3.52	NS
Workload Level	[W]	3	0.095	0.032	W x S(G)	4.20*	0.007
Interactions							
T x G		3	0.110	0.037	T x S(G)	0.62	NS
T x D		1	0.067	0.067	T x D x S(G)	3.64	NS
T x W		3	0.270	0.090	T x W x S(G)	10.60***	0.025
G x D		3	0.359	0.120	D x S(G)	3.37	NS
G x W		9	0.082	0.009	W x S(G)	1.21	NS
D x W		3	0.374	0.125	D x W x S(G)	15.21***	0.035
T x G x D		3	0.049	0.016	T x D x S(G)	0.89	NS
T x G x W		9	0.054	0.006	T x W x S(G)	0.71	NS
T x D x W		3	0.127	0.042	T x D x W x S(G)	5.62**	0.011
G x D x W		9	0.136	0.015	D x W x S(G)	1.84	NS
T x G x D x W		9	0.158	0.018	T x D x W x S(G)	2.32*	0.009
Error Terms							
S(G)		12	4.138	0.345			
T x S(G)		12	0.708	0.059			
D x S(G)		12	0.426	0.035			
W x S(G)		36	0.273	0.008			
T x D x S(G)		12	0.222	0.018			
T x W x S(G)		36	0.306	0.009			
D x W x S(G)		36	0.295	0.008			
T x D x W x S(G)		36	0.271	0.008			
Total		255	9.906				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E19. Five Factor ANOVA for Ellipticity Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.00420	0.00420	T x S(G)	3.59	NS
Group	[G]	3	0.00322	0.00107	S(G)	1.05	NS
#Subject(Group)	[S(G)]	12	0.04888	0.00407		0.11	
Time Day	[D]	1	0.00214	0.00214	D x S(G)	3.23	NS
Workload Level	[W]	3	0.41644	0.13881	W x S(G)	8.57***	0.060
Interactions							
T x G		3	0.00234	0.00078	T x S(G)	0.67	NS
T x D		1	0.00209	0.00209	T x D x S(G)	3.82	NS
T x W		3	0.44412	0.14804	T x W x S(G)	6.74**	0.061
G x D		3	0.00024	0.00008	D x S(G)	0.12	NS
G x W		9	0.17550	0.01950	W x S(G)	1.20	NS
D x W		3	1.22892	0.40964	D x W x S(G)	19.56***	0.189
T x G x D		3	0.00354	0.00118	T x D x S(G)	2.16	NS
T x G x W		9	0.11865	0.01318	T x W x S(G)	0.60	NS
T x D x W		3	0.17859	0.05953	T x D x W x S(G)	3.11*	0.020
G x D x W		9	0.23307	0.02590	D x W x S(G)	1.24	NS
T x G x D x W		9	0.47080	0.05231	T x D x W x S(G)	2.73*	0.048
Error Terms							
S(G)		12	0.01222	0.00102			
T x S(G)		12	0.01403	0.00117			
D x S(G)		12	0.00793	0.00066			
W x S(G)		36	0.58321	0.01620			
T x D x S(G)		12	0.00655	0.00055			
T x W x S(G)		36	0.79043	0.02196			
D x W x S(G)		36	0.75388	0.02094			
T x D x W x S(G)		36	0.68905	0.01914			
Total		255	6.14115				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E20. Five Factor ANOVA for Fraction Velocity Fixation Gate

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.000022	0.000022	T x S(G)	0.00	NS
Group	[G]	3	1.076850	0.358950	S(G)	0.87	NS
#Subject(Group)	[S(G)]	12	19.745390	1.645449		69.51***	
Time Day	[D]	1	0.191110	0.191110	D x S(G)	5.12*	0.014
Workload Level	[W]	3	0.094544	0.031515	W x S(G)	3.49*	0.006
Interactions							
T x G		3	0.107587	0.035862	T x S(G)	0.69	NS
T x D		1	0.124923	0.124923	T x D x S(G)	12.73**	0.011
T x W		3	0.420895	0.140298	T x W x S(G)	15.11***	0.036
G x D		3	0.170446	0.056815	D x S(G)	1.52	NS
G x W		9	0.065319	0.007258	W x S(G)	0.80	NS
D x W		3	0.501820	0.167273	D x W x S(G)	17.82***	0.044
T x G x D		3	0.047891	0.015964	T x D x S(G)	1.63	NS
T x G x W		9	0.094029	0.010448	T x W x S(G)	1.13	NS
T x D x W		3	0.165552	0.055184	T x D x W x S(G)	5.45**	0.012
G x D x W		9	0.074733	0.008304	D x W x S(G)	0.88	NS
T x G x D x W		9	0.199527	0.022170	T x D x W x S(G)	2.19*	0.010
Error Terms							
S(G)		12	4.936348	0.411362			
T x S(G)		12	0.622876	0.051906			
D x S(G)		12	0.447775	0.037315			
W x S(G)		36	0.324716	0.009020			
T x D x S(G)		12	0.117747	0.009812			
T x W x S(G)		36	0.334162	0.009282			
D x W x S(G)		36	0.337871	0.009385			
T x D x W x S(G)		36	0.364247	0.010118			
Total		255	10.820992				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E21. Five Factor ANOVA for Fraction Angle Fixation Gate

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.458	0.458	T x S(G)	9.42**	0.041
Group	[G]	3	0.268	0.089	S(G)	0.22	NS
#Subject(Group)	[S(G)]	12	19.334	1.611		69.49***	
Time Day	[D]	1	0.160	0.160	D x S(G)	5.61*	0.013
Workload Level	[W]	3	0.124	0.041	W x S(G)	5.55**	0.010
Interactions							
T x G		3	0.077	0.026	T x S(G)	0.52	NS
T x D		1	0.069	0.069	T x D x S(G)	3.21	NS
T x W		3	0.331	0.110	T x W x S(G)	10.65***	0.030
G x D		3	0.182	0.061	D x S(G)	2.13	NS
G x W		9	0.149	0.017	W x S(G)	2.23*	0.008
D x W		3	0.330	0.110	D x W x S(G)	14.03***	0.031
T x G x D		3	0.011	0.004	T x D x S(G)	0.17	NS
T x G x W		9	0.070	0.008	T x W x S(G)	0.75	NS
T x D x W		3	0.134	0.045	T x D x W x S(G)	4.50**	0.010
G x D x W		9	0.102	0.011	D x W x S(G)	1.45	NS
T x G x D x W		9	0.156	0.017	T x D x W x S(G)	1.75	NS
Error Terms							
S(G)		12	4.833	0.403			
T x S(G)		12	0.584	0.049			
D x S(G)		12	0.342	0.029			
W x S(G)		36	0.268	0.007			
T x D x S(G)		12	0.257	0.021			
T x W x S(G)		36	0.373	0.010			
D x W x S(G)		36	0.282	0.008			
T x D x W x S(G)		36	0.356	0.010			
Total		255	9.915				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E22. Five Factor ANOVA for Fraction Dual Fixation Gate

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.47322	0.47322	T x S(G)	78.81***	0.205
Group	[G]	3	0.35184	0.11728	S(G)	3.07	NS
#Subject(Group)	[S(G)]	12	1.83498	0.15292		21.22***	
Time Day	[D]	1	0.00085	0.00085	D x S(G)	0.15	NS
Workload Level	[W]	3	0.16349	0.05450	W x S(G)	19.92***	0.068
Interactions							
T x G		3	0.06769	0.02256	T x S(G)	3.76*	0.022
T x D		1	0.00563	0.00563	T x D x S(G)	0.99	NS
T x W		3	0.00674	0.00225	T x W x S(G)	0.93	NS
G x D		3	0.03518	0.01173	D x S(G)	2.06	NS
G x W		9	0.03449	0.00383	W x S(G)	1.40	NS
D x W		3	0.05393	0.01798	D x W x S(G)	8.21***	0.021
T x G x D		3	0.03384	0.01128	T x D x S(G)	1.97	NS
T x G x W		9	0.02000	0.00222	T x W x S(G)	0.92	NS
T x D x W		3	0.00172	0.00057	T x D x W x S(G)	0.40	NS
G x D x W		9	0.01522	0.00169	D x W x S(G)	0.77	NS
T x G x D x W		9	0.02947	0.00327	T x D x W x S(G)	2.27*	0.007
Error Terms							
S(G)		12	0.45875	0.03823			
T x S(G)		12	0.07205	0.00600			
D x S(G)		12	0.06832	0.00569			
W x S(G)		36	0.09847	0.00274			
T x D x S(G)		12	0.06858	0.00572			
T x W x S(G)		36	0.08727	0.00242			
D x W x S(G)		36	0.07887	0.00219			
T x D x W x S(G)		36	0.05203	0.00145			
Total		255	2.27765				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E23. Five Factor ANOVA for Percent Transition Matrix Symmetric

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	32.91	32.91	T x S(G)	0.74	NS
Group	[G]	3	1097.59	365.86	S(G)	1.97	NS
#Subject(Group)	[S(G)]	12	8914.71	742.89		17.57***	
Time Day	[D]	1	10.79	10.79	D x S(G)	0.25	NS
Workload Level	[W]	3	142.25	47.42	W x S(G)	2.31	NS
Interactions							
T x G		3	11.30	3.77	T x S(G)	0.08	NS
T x D		1	29.40	29.40	T x D x S(G)	2.01	NS
T x W		3	190.58	63.53	T x W x S(G)	4.15*	0.015
G x D		3	152.55	50.85	D x S(G)	1.18	NS
G x W		9	38.17	4.24	W x S(G)	0.21	NS
D x W		3	867.13	289.04	D x W x S(G)	22.26***	0.088
T x G x D		3	232.04	77.35	T x D x S(G)	5.29*	0.020
T x G x W		9	134.59	14.95	T x W x S(G)	0.98	NS
T x D x W		3	269.13	89.71	T x D x W x S(G)	7.18***	0.025
G x D x W		9	161.68	17.96	D x W x S(G)	1.38	NS
T x G x D x W		9	350.38	38.93	T x D x W x S(G)	3.12**	0.025
Error Terms							
S(G)		12	2228.68	185.72			
T x S(G)		12	535.60	44.63			
D x S(G)		12	519.22	43.27			
W x S(G)		36	738.00	20.50			
T x D x S(G)		12	175.55	14.63			
T x W x S(G)		36	551.33	15.31			
D x W x S(G)		36	467.54	12.99			
T x D x W x S(G)		36	449.78	12.49			
Total		255	9386.16				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E24. Five Factor ANOVA for Percent Transition Matrix Repeat

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	117.17	117.17	T x S(G)	0.44	NS
Group	[G]	3	5624.13	1874.71	S(G)	3.07	NS
#Subject(Group)	[S(G)]	12	29334.34	2444.53		25.21***	
Time Day	[D]	1	92.39	92.39	D x S(G)	0.34	NS
Workload Level	[W]	3	1571.72	523.91	W x S(G)	10.71***	0.041
Interactions							
T x G		3	1305.29	435.10	T x S(G)	1.62	NS
T x D		1	93.80	93.80	T x D x S(G)	1.41	NS
T x W		3	528.96	176.32	T x W x S(G)	4.56**	0.012
G x D		3	1420.57	473.52	D x S(G)	1.76	NS
G x W		9	299.37	33.26	W x S(G)	0.68	NS
D x W		3	844.26	281.42	D x W x S(G)	5.18**	0.019
T x G x D		3	326.88	108.96	T x D x S(G)	1.64	NS
T x G x W		9	385.85	42.87	T x W x S(G)	1.11	NS
T x D x W		3	422.00	140.67	T x D x W x S(G)	3.21*	0.008
G x D x W		9	516.82	57.42	D x W x S(G)	1.06	NS
T x G x D x W		9	244.90	27.21	T x D x W x S(G)	0.62	NS
Error Terms							
S(G)		12	7333.59	611.13			
T x S(G)		12	3224.00	268.67			
D x S(G)		12	3222.44	268.54			
W x S(G)		36	1761.55	48.93			
T x D x S(G)		12	795.97	66.33			
T x W x S(G)		36	1392.75	38.69			
D x W x S(G)		36	1956.62	54.35			
T x D x W x S(G)		36	1576.82	43.80			
Total		255	35057.85				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E25. Five Factor ANOVA for Percent Transition Matrix Useful

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.83	0.83	T x S(G)	0.00	NS
Group	[G]	3	7963.71	2654.57	S(G)	3.33	NS
#Subject(Group)	[S(G)]	12	38215.82	3184.65		23.12***	
Time Day	[D]	1	20.55	20.55	D x S(G)	0.03	NS
Workload Level	[W]	3	4020.20	1340.07	W x S(G)	20.43***	0.076
Interactions							
T x G		3	1163.33	387.78	T x S(G)	1.62	NS
T x D		1	21.89	21.89	T x D x S(G)	0.46	NS
T x W		3	155.58	51.86	T x W x S(G)	1.02	NS
G x D		3	1855.52	618.51	D x S(G)	0.88	NS
G x W		9	929.78	103.31	W x S(G)	1.58	NS
D x W		3	1360.16	453.39	D x W x S(G)	7.02***	0.023
T x G x D		3	1069.49	356.50	T x D x S(G)	7.43**	0.018
T x G x W		9	685.06	76.12	T x W x S(G)	1.49	NS
T x D x W		3	80.30	26.77	T x D x W x S(G)	0.47	NS
G x D x W		9	520.71	57.86	D x W x S(G)	0.90	NS
T x G x D x W		9	373.60	41.51	T x D x W x S(G)	0.73	NS
Error Terms							
S(G)		12	9553.96	796.16			
T x S(G)		12	2872.27	239.36			
D x S(G)		12	8471.24	705.94			
W x S(G)		36	2360.98	65.58			
T x D x S(G)		12	576.03	48.00			
T x W x S(G)		36	1839.28	51.09			
D x W x S(G)		36	2324.14	64.56			
T x D x W x S(G)		36	2052.10	57.00			
Total		255	50270.67				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E26. Five Factor ANOVA for Short Fixation

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	287.94	287.94	T x S(G)	4.92*	0.018
Group	[G]	3	218.49	72.83	S(G)	0.21	NS
#Subject(Group)	[S(G)]	12	16594.00	1382.83		24.90***	
Time Day	[D]	1	55.63	55.63	D x S(G)	2.93	NS
Workload Level	[W]	3	175.01	58.34	W x S(G)	2.16	NS
Interactions							
T x G		3	251.38	83.79	T x S(G)	1.43	NS
T x D		1	40.51	40.51	T x D x S(G)	1.79	NS
T x W		3	454.64	151.55	T x W x S(G)	5.29**	0.029
G x D		3	99.58	33.19	D x S(G)	1.75	NS
G x W		9	264.64	29.40	W x S(G)	1.09	NS
D x W		3	202.43	67.48	D x W x S(G)	2.55	NS
T x G x D		3	221.73	73.91	T x D x S(G)	3.27	NS
T x G x W		9	223.00	24.78	T x W x S(G)	0.86	NS
T x D x W		3	423.50	141.17	T x D x W x S(G)	6.77**	0.028
G x D x W		9	247.83	27.54	D x W x S(G)	1.04	NS
T x G x D x W		9	533.89	59.32	T x D x W x S(G)	2.84*	0.027
Error Terms							
S(G)		12	4148.50	345.71			
T x S(G)		12	702.11	58.51			
D x S(G)		12	227.94	19.00			
W x S(G)		36	970.59	26.96			
T x D x S(G)		12	271.50	22.62			
T x W x S(G)		36	1031.65	28.66			
D x W x S(G)		36	954.22	26.51			
T x D x W x S(G)		36	751.00	20.86			
Total		255	12757.72				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E27. Five Factor ANOVA for Viewing Cycles

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	68.67	68.67	T x S(G)	4.46	NS
Group	[G]	3	233.39	77.80	S(G)	1.22	NS
#Subject(Group)	[S(G)]	12	3060.72	255.06		28.18***	
Time Day	[D]	1	5.82	5.82	D x S(G)	0.49	NS
Workload Level	[W]	3	109.23	36.41	W x S(G)	8.23***	0.034
Interactions							
T x G		3	16.67	5.56	T x S(G)	0.36	NS
T x D		1	16.04	16.04	T x D x S(G)	2.41	NS
T x W		3	32.71	10.90	T x W x S(G)	4.11*	0.009
G x D		3	17.11	5.70	D x S(G)	0.48	NS
G x W		9	37.86	4.21	W x S(G)	0.95	NS
D x W		3	280.76	93.59	D x W x S(G)	20.81***	0.095
T x G x D		3	103.05	34.35	T x D x S(G)	5.15*	0.030
T x G x W		9	46.00	5.11	T x W x S(G)	1.93	NS
T x D x W		3	96.03	32.01	T x D x W x S(G)	14.07***	0.032
G x D x W		9	42.67	4.74	D x W x S(G)	1.05	NS
T x G x D x W		9	24.72	2.75	T x D x W x S(G)	1.21	NS
Error Terms							
S(G)		12	765.18	63.77			
T x S(G)		12	184.67	15.39			
D x S(G)		12	141.47	11.79			
W x S(G)		36	159.24	4.42			
T x D x S(G)		12	79.96	6.66			
T x W x S(G)		36	95.53	2.65			
D x W x S(G)		36	161.91	4.50			
T x D x W x S(G)		36	81.90	2.28			
Total		255	2800.59				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E28. Five Factor ANOVA for Fixation Time

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.01607	0.01607	T x S(G)	5.97*	0.040
Group	[G]	3	0.00802	0.00267	S(G)	0.25	NS
#Subject(Group)	[S(G)]	12	0.50456	0.04205		44.15***	
Time Day	[D]	1	0.00218	0.00218	D x S(G)	2.58	NS
Workload Level	[W]	3	0.01399	0.00466	W x S(G)	11.39***	0.039
Interactions							
T x G		3	0.00389	0.00130	T x S(G)	0.48	NS
T x D		1	0.00047	0.00047	T x D x S(G)	0.69	NS
T x W		3	0.01013	0.00338	T x W x S(G)	7.38***	0.027
G x D		3	0.00240	0.00080	D x S(G)	0.95	NS
G x W		9	0.00723	0.00080	W x S(G)	1.96	NS
D x W		3	0.00843	0.00281	D x W x S(G)	6.69**	0.022
T x G x D		3	0.00086	0.00029	T x D x S(G)	0.42	NS
T x G x W		9	0.00560	0.00062	T x W x S(G)	1.36	NS
T x D x W		3	0.00546	0.00182	T x D x W x S(G)	5.33**	0.014
G x D x W		9	0.00267	0.00030	D x W x S(G)	0.71	NS
T x G x D x W		9	0.00539	0.00060	T x D x W x S(G)	1.75	NS
Error Terms							
S(G)		12	0.12614	0.01051			
T x S(G)		12	0.03228	0.00269			
D x S(G)		12	0.01015	0.00085			
W x S(G)		36	0.01474	0.00041			
T x D x S(G)		12	0.00824	0.00069			
T x W x S(G)		36	0.01647	0.00046			
D x W x S(G)		36	0.01510	0.00042			
T x D x W x S(G)		36	0.01229	0.00034			
Total		255	0.32822				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

\# Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E29. Five Factor ANOVA for Fixation Time Change

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	0.000340	0.000340	T x S(G)	8.93*	0.001
Group	[G]	3	0.000161	0.000054	S(G)	1.98	NS
#Subject(Group)	[S(G)]	12	0.001302	0.000109		0.06	
Time Day	[D]	1	0.000664	0.000664	D x S(G)	8.58*	0.002
Workload Level	[W]	3	0.020610	0.006870	W x S(G)	7.49***	0.068
Interactions							
T x G		3	0.000173	0.000058	T x S(G)	1.52	NS
T x D		1	0.000016	0.000016	T x D x S(G)	0.34	NS
T x W		3	0.016682	0.005561	T x W x S(G)	4.77**	0.050
G x D		3	0.000067	0.000022	D x S(G)	0.29	NS
G x W		9	0.014694	0.001633	W x S(G)	1.78	NS
D x W		3	0.018820	0.006273	D x W x S(G)	5.24**	0.058
T x G x D		3	0.000093	0.000031	T x D x S(G)	0.66	NS
T x G x W		9	0.013224	0.001469	T x W x S(G)	1.26	NS
T x D x W		3	0.007465	0.002488	T x D x W x S(G)	2.79	NS
G x D x W		9	0.004326	0.000481	D x W x S(G)	0.40	NS
T x G x D x W		9	0.012701	0.001411	T x D x W x S(G)	1.58	NS
Error Terms							
S(G)		12	0.000326	0.000027			
T x S(G)		12	0.000457	0.000038			
D x S(G)		12	0.000929	0.000077			
W x S(G)		36	0.033034	0.000918			
T x D x S(G)		12	0.000565	0.000047			
T x W x S(G)		36	0.041925	0.001165			
D x W x S(G)		36	0.043064	0.001196			
T x D x W x S(G)		36	0.032066	0.000891			
Total		255	0.262401				

* = $p<0.05$ ** = $p<0.01$ *** = $p<0.001$ NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E30. Five Factor ANOVA for Long Fixation

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	634.15	634.15	T x S(G)	7.31*	0.043
Group	[G]	3	640.82	213.61	S(G)	0.50	NS
#Subject(Group)	[S(G)]	12	20431.61	1702.63		54.33***	
Time Day	[D]	1	134.54	134.54	D x S(G)	2.55	NS
Workload Level	[W]	3	722.08	240.69	W x S(G)	16.30***	0.053
Interactions							
T x G		3	102.83	34.28	T x S(G)	0.40	NS
T x D		1	5.13	5.13	T x D x S(G)	0.46	NS
T x W		3	315.44	105.15	T x W x S(G)	7.81***	0.021
G x D		3	49.39	16.46	D x S(G)	0.31	NS
G x W		9	154.03	17.11	W x S(G)	1.16	NS
D x W		3	419.35	139.78	D x W x S(G)	9.18***	0.029
T x G x D		3	53.99	18.00	T x D x S(G)	1.60	NS
T x G x W		9	235.67	26.19	T x W x S(G)	1.94	NS
T x D x W		3	148.29	49.43	T x D x W x S(G)	3.70*	0.008
G x D x W		9	90.95	10.11	D x W x S(G)	0.66	NS
T x G x D x W		9	120.53	13.39	T x D x W x S(G)	1.00	NS
Error Terms							
S(G)		12	5107.90	425.66			
T x S(G)		12	1040.92	86.74			
D x S(G)		12	632.64	52.72			
W x S(G)		36	531.64	14.77			
T x D x S(G)		12	134.60	11.22			
T x W x S(G)		36	484.69	13.46			
D x W x S(G)		36	548.30	15.23			
T x D x W x S(G)		36	481.25	13.37			
Total		255	12789.13				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

Table E31. Five Factor ANOVA for Index of Engagement

Effect		df	SS	MS	Error Term	F	ω^2
Main Effects							
Task	[T]	1	1.059	1.059	T x S(G)	38.45***	0.071
Group	[G]	3	2.260	0.753	S(G)	1.12	NS
#Subject(Group)	[S(G)]	12	32.310	2.692		91.07***	
Time Day	[D]	1	0.011	0.011	D x S(G)	0.64	NS
Workload Level	[W]	3	0.183	0.061	W x S(G)	7.03***	0.011
Interactions							
T x G		3	0.063	0.021	T x S(G)	0.76	NS
T x D		1	0.027	0.027	T x D x S(G)	1.60	NS
T x W		3	0.045	0.015	T x W x S(G)	1.89	NS
G x D		3	0.035	0.012	D x S(G)	0.65	NS
G x W		9	0.183	0.020	W x S(G)	2.34*	0.007
D x W		3	0.111	0.037	D x W x S(G)	7.72***	0.007
T x G x D		3	0.034	0.011	T x D x S(G)	0.67	NS
T x G x W		9	0.137	0.015	T x W x S(G)	1.92	NS
T x D x W		3	0.076	0.025	T x D x W x S(G)	2.62	NS
G x D x W		9	0.139	0.015	D x W x S(G)	3.23**	0.007
T x G x D x W		9	0.106	0.012	T x D x W x S(G)	1.21	NS
Error Terms							
S(G)		12	8.077	0.673			
T x S(G)		12	0.331	0.028			
D x S(G)		12	0.214	0.018			
W x S(G)		36	0.313	0.009			
T x D x S(G)		12	0.204	0.017			
T x W x S(G)		36	0.286	0.008			
D x W x S(G)		36	0.172	0.005			
T x D x W x S(G)		36	0.349	0.010			
Total		255	14.415				

* = p<0.05 ** = p<0.01 *** = p<0.001 NS = not significant

Refers to an ANOVA using the variance of replications as the error term. All other tests were from an ANOVA using the mean of replications as the dependent variable.

VITA

Daniel J. Callan received his Bachelor of Science degree in Chemical Engineering from the University of Notre Dame in 1982. In 1987, he received his Master of Science degree in Mechanical Engineering from Boston University. In December, 1998 Daniel J. Callan graduated from the Pennsylvania State University with his Doctorate in Industrial Engineering, Human Factors.

Professionally, Dan is a Major in the United States Air Force. He began his career as a Weapon Specialist Officer flying RF-4 reconnaissance aircraft. Upon graduation from Test Pilot School at Edwards, AFB in 1990, he was program manager for the Advanced Tactical Air Reconnaissance System and flew the F-15E for flight test missions involving munitions separation. In 1994, Dan was accepted into AFIT to complete a Doctorate degree in Human Factors. He is stationed currently at Wright-Patterson AFB working in the Human Effectiveness Directorate as the Chief of the Information Analysis and Exploitation Branch.

Daniel J. Callan with his instructor, Associate Professor Joseph Goldberg, presented a paper in Derby, England at the 1996 European Conference in Eye Movement. The topic was Fitts Law applications in eye movement. In Chicago, IL, he presented "Eye Movement Relationships to Excessive Performance Error in Aviation" at the 1998 Human Factors and Ergonomic Society's 42nd Annual Meeting.

ABSTRACT
Psychophysiological Measures for Human Attention Lapses During Simulated Aircraft Operations
Daniel J. Callan
Ph.D.; December 1998
The Pennsylvania State University
Joseph H. Goldberg, Thesis Advisor

This study produced a range of aviation performance to which psychophysiological measures were correlated predicting performance decrements due to task overload and vigilance decrement. A high fidelity simulation of an instrument flight pattern produced multiple workload levels resulting in various levels of performance. Psychophysiological parameters including eye movements, EEG, and peripheral temperature were measured. Workload was varied and a secondary task was added to create realistic operational performance levels. Four groups of four subjects provided 64 data segments each during two, 2 hour simulation periods. Eight subjects were instrument rated and eight unrated. Eight subjects had commercial flight experience and eight had no commercial flight experience. Operationally relevant performance levels were based upon Air Traffic Control (ATC) and safety standards. Subjects' performance error was dangerous for 18 of 1024 segments and exceeded ATC standards on additional 193 segments. The Long Fixation parameter was sensitive enough to predict 83% of segments exceeding ATC performance error standards with a 15% false alarm rate.

Factors of workload, attentiveness, and cognitive processing capability affect performance; different psychophysiological parameters are needed to completely describe performance. Level of arousal reflected the "level of attention" for perception, processing, and response execution. The two best arousal parameters, Peripheral Temperature Change and Pupil Diameter Change, were the best performance predictors,

these parameters reflected performance decrements related to workload and other stressors. Performance decrements associated with nominal or low workloads were not detected. Saccade Time, Dual Fixation Gate, and seven other parameters related to task type showed great promise in providing real time feedback on workload levels and the type of task on which operators are engaged.

Elements of cognitive performance were described by the Long Fixation and Short Fixation parameters. A high frequency of Long Fixations was indicative of problem solving activity. A high frequency of Short Fixations was indicative of efficient processing. However, the efficiency was not related to only to workload since subjects used large numbers of short fixations when monitoring the simulation.